

# Type 1 AGN at low $z$ . I. Emission properties

Jonathan Stern\* and Ari Laor\*

*Department of Physics, Technion – Israel Institute of Technology, Haifa 32000, Israel*

27 February 2013

## ABSTRACT

We analyze the emission properties of a new sample of 3 579 type 1 AGN, selected from the SDSS DR7 based on the detection of broad  $H\alpha$  emission. The sample extends over a broad  $H\alpha$  luminosity  $L_{bH\alpha}$  of  $10^{40} - 10^{44}$  erg s $^{-1}$  and a broad  $H\alpha$  FWHM of 1 000 – 25 000 km s $^{-1}$ , which covers the range of black hole mass  $10^6 < M_{BH}/M_{\odot} < 10^{9.5}$  and luminosity in Eddington units  $10^{-3} < L/L_{Edd} < 1$ . We combine ROSAT, GALEX and 2MASS observations to form the SED from 2.2  $\mu$ m to 2 keV. We find the following: 1. The distribution of the  $H\alpha$  FWHM values is independent of luminosity. 2. The observed mean optical-UV SED is well matched by a fixed shape SED of luminous quasars, which scales linearly with  $L_{bH\alpha}$ , and a host galaxy contribution. 3. The host galaxy  $r$ -band (fibre) luminosity function follows well the luminosity function of inactive non-emission line galaxies (NEG), consistent with a fixed fraction of  $\sim 3\%$  of NEG hosting an AGN, regardless of the host luminosity. 4. The hosts of lower luminosity AGN have a mean  $z$  band luminosity and  $u - z$  colour which are identical to NEG with the same redshift distribution. With increasing  $L_{bH\alpha}$  the AGN hosts become bluer and less luminous than NEG. The implied increasing star formation rate with  $L_{bH\alpha}$  is consistent with the relation for SDSS type 2 AGN of similar bolometric luminosity. 5. The optical-UV SED of the more luminous AGN shows a small dispersion, consistent with dust reddening of a blue SED, as expected for thermal thin accretion disc emission. 6. There is a rather tight relation of  $\nu L_{\nu}(2\text{keV})$  and  $L_{bH\alpha}$ , which provides a useful probe for unobscured (true) type 2 AGN. 7. The primary parameter which drives the X-ray to UV emission ratio is the luminosity, rather than  $M_{BH}$  or  $L/L_{Edd}$ .

**Key words:**

## 1 INTRODUCTION

The first systematic study, based on optical spectroscopy, of a complete and well defined sample of Broad Line AGN was conducted by Boroson & Green (1992, hereafter BG92). It included 87  $z < 0.5$  AGN from the Bright Quasar Survey (BQS, Green et al. 1986), with spectra taken over the 4100Å–5900Å range, at a spectral resolution of  $\sim 700$ . This survey, and additional systematic studies at other wavelengths, led to major new understandings of the emission and absorption properties of AGN (BG92, and citations thereafter). The BQS survey was followed by the Large Bright Quasar Survey, which yielded spectra of 1055 quasars (Hewett et al. 1995), and by the 2dF QSO Redshift Survey which yielded spectra for 23 338 quasars (Croom et al. 2004). The next major step in AGN optical spectroscopy came with the Sloan Digital Sky Survey (SDSS) Quasar Catalog (latest release QCV, Schneider et al. 2010), which produced a sample of quasars a factor of 1000 larger, compared to BG92, with a factor of 10 more spectroscopic data per object (a factor of three larger wavelength coverage and a factor of three larger

spectral resolution). This sample is now a prime resource for studies of AGN (see Schneider et al. 2002, and citations thereafter). The QCV sample is mainly based on objects selected for spectroscopy due to their non-stellar colours, supplemented by SDSS spectra selected based on other surveys. The QCV sample includes luminous objects ( $M_i < -22.0$ ) which exhibit at least one emission line with FWHM  $> 1000$  km s $^{-1}$ .

Various studies used the SDSS to produce samples of lower luminosity AGN. At  $M_i > -22.0$  the AGN are generally dominated by the host light, which needs to be subtracted to detect the AGN continuum and broad line emission. Hao et al. (2005a), who used the SDSS second data release (DR2), subtracted the host light based on a principal component analysis method. They then modeled the  $H\alpha$  profile using one or two Gaussians, and derived the FWHM of the broader component. Their type 1 AGN sample includes a total of 1317 objects with FWHM  $> 1200$  km s $^{-1}$ , and was used to study the extension of the luminosity function of type 1 AGN to low luminosities. The later study of Vanden Berk et al. (2006), based on DR3, extended the earlier SDSS quasar sample (Schneider et al. 2005) to lower luminosity by removing the absolute magnitude criterion, which yielded 4 666 low-luminosity type 1 AGN, used to study the relation between the AGN luminosity and the host prop-

\* E-mail: stern@physics.technion.ac.il (JS);  
laor@physics.technion.ac.il (AL)

erties. The definition of a type 1 AGN in Vanden Berk et al. (2006), and also in QCV, requires the detection of at least one emission line with  $\text{FWHM} > 1000 \text{ km s}^{-1}$ . This criterion works well for luminous AGN, as the emission line width is dominated by the Broad Line Region (BLR) in type 1 AGN, and it excludes type 2 AGN where the emission lines are from the Narrow Line Region (NLR), where the  $\text{FWHM} < 1000 \text{ km s}^{-1}$ . However, as we show below lower luminosity type 1 AGN show an increasing relative contribution from the NLR, and would fail the  $\text{FWHM} > 1000 \text{ km s}^{-1}$  criterion. Thus, the Vanden Berk et al. (2006) sample becomes increasingly incomplete with decreasing AGN luminosity. The latest SDSS sample of low luminosity type 1 AGN is of Green & Ho (2007, hereafter GH07), which is based on DR4. This sample is based on the detection of a broad component for the  $\text{H}\alpha$  emission line in the  $6400\text{\AA} - 6700\text{\AA}$  wavelength region. It includes 8435 objects, and was used to study the AGN Black Hole mass ( $M_{\text{BH}}$ ) function. We note that broad  $\text{H}\alpha$  with  $\text{FWHM} > 10,000 \text{ km s}^{-1}$  are selected against in the GH07 sample. due to their restricted search region.

Here we present a new SDSS based sample of type 1 AGN which extends to low luminosity, with the purpose of providing a useful and well defined database which extends the QCV sample to low luminosity, using DR7 (Abazajian et al. 2009). We combine this data set with observations in the UV, IR, and X-ray, based on GALEX (Martin et al. 2005), 2MASS (Skrutskie et al. 2006) and ROSAT (Voges et al. 1999, 2000), in order to study the NIR to X-ray Spectral Energy Distribution (SED) of low luminosity AGN, and its dependence on various AGN properties. We also derive the mean properties of low luminosity AGN host galaxies, and compare them to the mean properties of inactive galaxies. In following papers we study some of the emission line properties of this sample.

The first study of the IR to X-ray SED of a complete and well defined sample of AGN was made by Sanders et al. (1989) for the BQS optically selected quasar sample. It showed the overall similarity of the SED of AGN, composed of an IR bump, an optical-UV bump, and an X-ray power-law continuum. A follow up study was made by Richards et al. (2006) for a small subsample of 259 quasars from the SDSS, which determined the wavelength dependence of the bolometric correction factor. However, little is known about the dependence of the AGN SED on luminosity,  $M_{\text{BH}}$ , and luminosity in Eddington units  $L/L_{\text{Edd}}$  (Vasudevan & Fabian 2009; Grupe et al. 2010). This dependence may provide important clues to the nature of the accretion process (e.g. Davis & Laor 2011).

Here we provide an order of magnitude larger sample of type 1 AGN with nearly complete detections in most bands. These allow us to form an unbiased SED of the sample which extends to low luminosity AGN, and to study its relation with various AGN properties. A significant limitation is that the SED of low luminosity AGN is often dominated by the host galaxy emission. High angular resolution, not available here, is required to isolate the low luminosity AGN emission from the host galaxy light (e.g. Maoz et al. 2005).

In addition, we explore how broad the broad  $\text{H}\alpha$  gets. Laor (2003) noted that type 1 AGN with BLR  $\text{FWHM} > 25,000 \text{ km s}^{-1}$  are extremely rare. Various theoretical models of the BLR and the ionizing continuum predict an upper limit on the FWHM, beyond which the BLR does not exist (Nicastro 2000, Elitzur and Shlosman 2006, Laor & Davis 2011). The models differ in the dependence of the maximal FWHM on luminosity. Here, we use the new sample to explore the FWHM distribution and its dependence on luminosity.

The paper is organized as follows. In §2.1 – §2.5 we describe our type 1 sample selection. In §2.6 we explore the selection ef-

fects which arise from our detection procedure, and in §2.7 we add photometry from IR, UV, and X-ray surveys. In §3.1 the distribution of the type 1 sample in various emission properties is explored, and compared to previous AGN samples, and to samples of inactive galaxies. In §3.2 – §3.4, we decompose the mean observed NIR – UV SED into the net-AGN and host SEDs. The implied host properties are examined in §3.5. In §3.6 we explore the relation between  $\nu L_{\nu}(\text{X-ray})$  and  $L_{\text{bH}\alpha}$ . The origin of the SED dispersion around the mean SED is explored in §3.7. In §3.8 – §3.9 we investigate the dependence of the AGN SED on  $M_{\text{BH}}$  and  $L/L_{\text{Edd}}$ . We discuss our results in §4, and summarize them in §5.

Throughout the paper, we assume a FRW cosmology with  $\Omega = 0.3$ ,  $\Lambda = 0.7$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2 SAMPLE CREATION

### 2.1 parent sample

Our purpose is to derive a complete and well defined sample of low luminosity broad line AGN, based on the SDSS DR7 (Abazajian et al. 2009). The SDSS obtained imaging of a quarter of the sky in five bands (*ugriz*; Fukugita et al. 1996) with a CCD mosaic camera (Gunn et al. 1998) coupled to a 2.5-m telescope. The 95% completeness limit in the *r* band is 22.2 mag. Spectroscopic follow-ups of the photometric catalog were performed using spectrographs fed by  $3''$ -aperture fibres. Targets were selected for spectroscopic followup if one of the following criteria applied to the photometry: an extended morphology, which creates the  $\sim 10^6$  objects of the main galaxy sample (Strauss et al. 2002); non-stellar colours or a nearby FIRST (Becker et al. 1995) source, which is the base of the  $\sim 10^5$  objects of QCV. Based on the availability of fibers, targets with a nearby ROSAT (Voges et al. 1999, 2000) source might have also been selected for spectroscopy. The spectrographs cover the wavelength range between  $3800\text{\AA}$  and  $9200\text{\AA}$  at a resolution of  $\sim 150 \text{ km s}^{-1}$ , with a typical S/N of  $10 \text{ pixel}^{-1}$  for a galaxy near the flux limit. The flux of the spectra is calibrated by matching the spectra of simultaneously observed standard stars to their PSF magnitude (Adelman-McCarthy et al. 2008). On each spectrum, the SDSS pipeline measures the emission lines, and then determines the redshift  $z$  and the classification of the spectrum, according to the locations of emission lines, or according to correlations with template spectra.

We apply several initial filters to the DR7 database:

- (i) The spectrum is classified as non-stellar.
- (ii)  $z < 0.31$ , to ensure the continuum level can be measured at least  $20,000 \text{ km s}^{-1}$  to the red away from  $\text{H}\alpha$ , so that very broad  $\text{H}\alpha$  lines with  $\text{FWHM} \lesssim 25,000 \text{ km s}^{-1}$  are detectable.
- (iii)  $z > 0.005$ , to ensure reliable distance estimates based on the assumed  $H_0$ , without significant errors due to peculiar motions.
- (iv) A S/N  $> 10$  in the spectrum<sup>1</sup>, to improve the accuracy of the NLR/BLR decomposition.
- (v) A percentage of pixels near  $\text{H}\alpha$  indicated as ‘bad’ in the pixel mask (Stoughton et al. 2002) which is no more than 0%, 20% or 50%, depending on the distance from  $\text{H}\alpha$  (see appendix A1). The main source of bad pixels is poor sky subtraction, which degrades the spectrum mainly at  $8000\text{\AA} < \lambda$ .

<sup>1</sup> Based on median S/N of spectrum pixels covered by the *g*, *r*, *i* bands (available in the SDSS database).

A total of 232 837 spectra, out of the 1.6 million spectra in DR7, pass these five criteria. We dub them as the ‘parent sample’. The spectra are then corrected for foreground dust, using the maps of Schlegel et al. (1998) and the extinction law of Cardelli et al. (1989). We now need to determine which of these objects host a broad line AGN.

## 2.2 Host subtraction

The rather large  $3''$  fibre aperture of the SDSS spectra implies that a significant fraction of the host light is typically included, which needs to be removed in order to identify low luminosity AGN. As discussed below (§2.3), the AGN are identified through the detection of a broad  $H\alpha$  component. To detect this component we need to subtract the broad stellar absorption features, and derive a smooth continuum. In particular, in the vicinity of the  $H\alpha$  feature, there are flanking TiO bands (Bica & Alloin 1986, and see appendix A2), which need to be properly subtracted off.

To roughly model the absorption features, we follow an approach similar in principle to that described in Hao et al. (2005a). We utilize the galaxy eigenspectra (ESa) presented in Yip et al. (2004; hereafter Yip04), which were derived from a principle component analysis of SDSS galaxies. A linear sum of the first three ESa and a power law component with  $L_\lambda \propto \lambda^{-1.5}$ , representing the AGN continuum, is fit to wavelength regions which lack the strongest galaxy and AGN emission lines (see appendix A2). Note that the four components of the fit are non-orthogonal, in contrast with the original ESa of Yip04.

In practice, light emitted by hot stars near  $H\alpha$  is degenerate with the power-law component, as can be seen in the third ES of Yip04 (fig. 5 there). This degeneracy produces an unrealistic result that the mean fraction of continuum emission emitted by AGN in all galaxies is  $\approx 5\%$ . In the more thorough PCA decomposition performed by Vanden Berk et al. (2006), in which they use also eigenvectors derived from a sample of quasars, they found that host-AGN separation is unreliable if the contribution of either component is  $< 10\%$ . This problem probably also arises from the non-orthogonality of the base spectra, as we find here. Note that if one excludes objects in which one of the components is weak, a correlation between the host and AGN luminosity is introduced. To avoid these complications and uncertainties, we use our PCA results only to subtract the stellar features and search for a broad  $H\alpha$ , and not to derive the host and AGN emission.

The linear sum of the three ESa is subtracted from the SDSS spectra in the wavelength ranges  $4800\text{\AA}$ – $5250\text{\AA}$  and  $6125\text{\AA}$ – $7000\text{\AA}$ , to derive a featureless continuum near the  $H\alpha$ ,  $H\beta$  and [O III] lines. Note that the Yip04 ESa include narrow emission lines from H II regions, with typical FWHM of  $\sim 250 \text{ km s}^{-1}$ , which should be avoided. We interpolate over all the strong narrow features in the ESa prior to their subtraction (note the narrow lines do not participate in the fit). Since  $H\beta$  in ES 1 and 2, and  $H\alpha$  in ES 1, reside in wider stellar absorption features (FWHM  $\sim 1300 \text{ km s}^{-1}$ ), there may be additional narrower stellar absorption features, which are lost by the interpolation, and not properly subtracted off with the other stellar absorption features. The possible amplitude of this effect is small, and is discussed in following papers on the narrow emission line properties.

In 1% of the spectra in the parent sample only poor fits (reduced  $\chi^2 > 2$ ) are obtained. The algorithm described below might falsely detect a broad  $H\alpha$  feature when the fit is poor, and such false detections are filtered in the final stage of the pipeline (§2.5).

## 2.3 Candidate broad $H\alpha$ objects

A broad emission line is a unique signature of AGN emission<sup>2</sup>. We search for a broad  $H\alpha$  line, as  $H\alpha$  is the strongest AGN line in the visible region. Identifying AGN through the detection of a broad  $H\alpha$  line is easier than the detection of a power-law component, as the average flux density of the broad  $H\alpha$  feature is  $\sim 5$  times larger than the power-law continuum flux density (Fig. 1). Also, as discussed above, the power-law component is degenerate with the continuum of young stars. To identify a broad  $H\alpha$  line we search for excess flux above the continuum in the vicinity of  $H\alpha$ , which cannot be attributed to be part of the narrow emission line profiles.

After subtracting the stellar emission, the featureless continuum is derived by linearly interpolating the mean continuum level at the continuum windows at  $6125\text{\AA}$ – $6250\text{\AA}$  and  $6880\text{\AA}$ – $7000\text{\AA}$ , which are  $\sim 300\text{\AA}$ , or  $\sim 14000 \text{ km s}^{-1}$  away from  $H\alpha$ . The continuum is subtracted, and a broad  $H\alpha$  is looked for in the residual spectrum. The residual flux at  $6250\text{\AA}$ – $6880\text{\AA}$  is summed, excluding regions 20-pixel wide ( $1380 \text{ km s}^{-1}$ ) centred on the [O I], [S II], [N II] and  $H\alpha$  narrow emission lines. Also, we exclude excess flux detected away from  $H\alpha$ , with a region at least ten pixels wide with null mean residual flux, as it is implausible that such distinct emission features originate from a broad  $H\alpha$  emission feature. Bad pixels are interpolated over. The total excess flux is denoted by  $\Delta F$ .

The significance of  $\Delta F$  is assessed by comparing it to  $\sigma_{\text{data}}$ , the dispersion in the flux density at  $6125\text{\AA}$ – $6250\text{\AA}$  and  $6880\text{\AA}$ – $7000\text{\AA}$ , which incorporates both the S/N of the spectrum and residual stellar features which were not fully subtracted in §2.2. A total of 6986 (3%) objects with  $\Delta F/\sigma_{\text{data}} > 2.5$ , from the parent sample of 232 837 objects, are found as potential broad-line AGN, and were passed on to the next stage of the pipeline. A depiction of this process, and the reasoning behind the 2.5 threshold, appear in appendix A3.

## 2.4 Broad $H\alpha$ and narrow lines fit

The purpose of this stage is to fit the excess flux near  $H\alpha$  by a broad line profile. We measure the profiles of various forbidden lines, [O III]  $\lambda 5007$ , [O I]  $\lambda\lambda 6300, 6363$ , [N II]  $\lambda\lambda 6548, 6583$  and [S II]  $\lambda\lambda 6716, 6730$ , which provide the narrow lines profile, and help us separate the broad and narrow components of  $H\alpha$ . We use 4<sup>th</sup>-order Gauss-Hermite functions (GHs; van der Marel & Franx 1993) to fit the narrow line profiles, and an up to 10<sup>th</sup>-order GH to fit the broad  $H\alpha$  profile, which can have a complex shape, such as a double peak. The Levenberg-Marquardt best-fit algorithm (Press et al. 1992) is used in all fits. The fit is used to measure the luminosities and mean velocities of all the lines mentioned above, and the higher moments are measured only for the broad  $H\alpha$  and for [O III].

An overview of the fitting process is given below, and further details appear in appendix A4. The narrow line results are used in following papers, so a more elaborate analysis of their fit procedure and its implications is given there.

The wavelength region near [O III],  $4967\text{\AA}$ – $5250\text{\AA}$ , is fit with three components: an iron template derived from observations of 1 Zw 1 (kindly provided by T. Boroson), a GH fit for the [O III] line, and a continuum component. The iron template can broaden and

<sup>2</sup> Excluding the rare examples of broad  $H\alpha$  due to supernovae (see figures 13 & 14 in Filippenko 1997). Only a handful of such objects reached the final sample (§2.5).

shift during the fit. Following that, all the lines between 6250Å–6880Å noted above are fit, using a 4<sup>th</sup>-order GH for the broad H $\alpha$ . To lower the degrees of freedom, the widths and higher moments of all the narrow lines in this region are kept equal. In the cases where all these narrow lines are too weak to significantly constrain their fit (14% of final sample, see §2.5), we tie their widths and higher moments to that of [O III], which is detected in 99% of the objects in the final sample. Then, the broad H $\alpha$  is fit with increasingly higher order GHs, until an acceptable fit is achieved. Finally, the narrow H $\beta$  line is also fit, keeping its centroid, width and shape equal to that of the narrow H $\alpha$ . Fitting the broad H $\beta$  profile is beyond the scope of this paper.

Following the above process, 3% of the objects have an unacceptable reduced  $\chi^2 > 2$  in the 6250Å–6880Å region. The treatment of these objects is detailed in appendix A4.

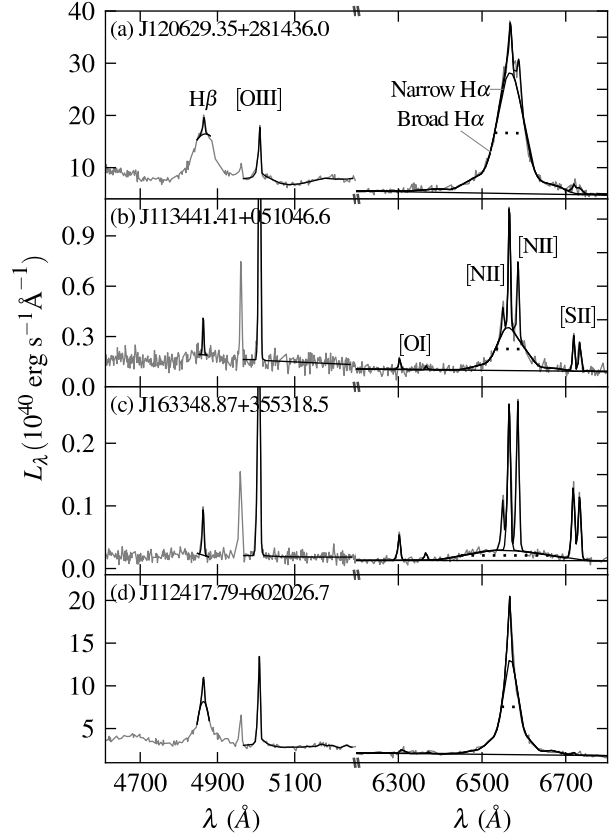
Figure 1 shows various examples of fits. The three upper panels show objects with a range of broad to narrow H $\alpha$  peak flux density ratio. The upper panel shows a ‘standard’ Seyfert 1.0, where the broad H $\alpha$  dominates, with a flux density ratio at the line core of  $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 3$ . Note the extended wings of the broad H $\alpha$  profile, beyond a simple Gaussian, which are fit by higher order GH components. The second panel shows a Seyfert 1.5 (Osterbrock 1981), where the narrow H $\alpha$  peak is higher than the broad H $\alpha$  peak  $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 0.5$ . The FWHM of the total H $\alpha$  profile is dominated by the narrow component. Such an object is therefore excluded from QCV and the Vanden Berk et al. (2006) sample of broad line AGN, although it is clearly a broad line AGN. The third panel shows a Seyfert 1.9 ( $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 0.05$ ), where a very broad H $\alpha$  is still detectable. The broad H $\beta$  is unobservable, as expected. Such an object will be defined as a Seyfert 2, if only H $\beta$  is available, although the EW of H $\alpha$  (before the host subtraction) is 28Å, which may indicate this is actually an unobscured broad line AGN. A full analysis of narrow to broad flux ratios and their dependence on AGN characteristics is deferred to a following paper.

The bottom panel is an example of an object with no clear transition between the Balmer narrow and broad line profiles. Additionally, all narrow lines near H $\alpha$  are undetectable. In such objects, we use the [O III]  $\lambda$ 5007 profile as a template profile for the narrow Balmer lines.

## 2.5 Final selection criteria

The purpose of this stage is to determine if the excess flux found above in 6986 objects is produced by a broad H $\alpha$ . As the  $\Delta F/\sigma_{\text{data}} = 2.5$  threshold is chosen to minimize the rejection of broad line AGN (appendix A3), many objects with larger  $\Delta F/\sigma_{\text{data}}$  lack a clear broad H $\alpha$  feature. To filter out such cases, we employ the additional following selection criteria:

- (i) A line width  $1 < \Delta v_{1000} < 25$ , where  $\Delta v_{1000} = \Delta v/1000$  and  $\Delta v$  is the H $\alpha$  FWHM in km s<sup>-1</sup>. The upper limit filters out cases where a very broad GH component essentially fits the continuum. The lower limit filters out a very narrow GH component which fits the narrow H $\alpha$  profile.
- (ii) A normalized excess flux density at the line centre  $F_{\lambda, \text{bH}\alpha}/\epsilon_{\lambda} > 2$ , where  $\epsilon_{\lambda}$  is the SDSS flux density error. This ensures significant flux density excess at the expected centre of H $\alpha$ , in addition to the above requirement of significant total flux. It filters out significant excess flux placed away from the position of H $\alpha$ . Also, it excludes broad and very shallow emission features produced when the continuum is not properly placed (e.g. due to residual stellar absorption features, §2.2).



**Figure 1.** Examples of the range of H $\alpha$  and H $\beta$  profiles observed in the host-subtracted spectra. The black lines are the fits to the broad H $\alpha$  (8<sup>th</sup>–10<sup>th</sup> order GH functions) and to the narrow lines (4<sup>th</sup>-order GHs). Dotted lines mark the FWHM of the Broad H $\alpha$ . (a) Seyfert 1.0,  $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 3$ : note the extended wings of the BL profile compared to a Gaussian, fit by high-order GH components. (b) Seyfert 1.5,  $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 0.5$ : the FWHM of the total H $\alpha$  profile is dominated by the narrow H $\alpha$ , therefore this object is excluded from the SDSS quasar catalog and the Vanden Berk et al. (2006) sample. (c) Seyfert 1.9,  $F_{\lambda, \text{bH}\alpha}/F_{\lambda, \text{nH}\alpha} \simeq 0.05$ : such an object will be defined as a Seyfert 2 if only H $\beta$  is available, despite an  $\text{EW}_{\text{bH}\alpha} = 28\text{\AA}$ , which may indicate this is an unobscured broad line AGN. (d) A type 1 AGN with no clear transition in the profile between  $F_{\lambda, \text{bH}\alpha}$  and  $F_{\lambda, \text{nH}\alpha}$ . The Balmer NL profiles are based on [O III], since [N II], [S II] and [O I] are undetectable.

- (iii) A revised excess flux criterion  $\Delta F'/\sigma_{\text{data}} > 4$ , rather than  $\Delta F/\sigma_{\text{data}} > 2.5$  used earlier (§2.3), where  $\Delta F'$  is obtained from the broad H $\alpha$  profile fit. Flux at  $\lambda$  less than  $3\sigma$  away from the peak positions of the narrow lines is excluded in the summation of  $\Delta F'$  (using the  $\sigma$  and peak fit to the narrow lines). This filters out objects in which the original excess flux may come from a broad base of the forbidden lines, rather than from the BLR.

A total of 3602 objects with acceptable fits (reduced  $\chi^2 < 2$ ) pass these criteria. The selection effects produced by the above criteria are discussed below. To verify the reality of the broad H $\alpha$  detection, we also inspected by eye all the 3602 spectra, and excluded further 93 objects where there was no clear BLR, i.e. a false-positives rate of  $\sim 3\%$  for our broad H $\alpha$  detection algorithm. Also, 80 objects with high  $\chi^2$  have an acceptable fit, as described in appendix A4. Finally, we inspected by eye the images of all the  $z < 0.03$  objects and removed three objects which were offset from

the centre of their host galaxies. The size of the final sample, hereafter the ‘T1 sample’, is 3 579 objects.

Table 1 lists the broad  $H\alpha$  luminosity ( $L_{bH\alpha}$ ), and  $\Delta v_{1000}$  for each object in the T1 sample. Our rejection criteria may have been too restrictive, and potentially excluded objects with an unusual broad line profile. We therefore provide in appendix B a table which lists the other 3 407 excess flux objects that were rejected from the T1 sample, along with the reason for rejection. This list may be useful for future studies.

## 2.6 Selection effects

Since the T1 sample is selected based on the detection of a broad  $H\alpha$ , it is important to understand how the sample selection criteria (§2.3 and §2.5) affect the range of detectable line widths, as a function of line flux.

Figure 2 presents three panels with the distribution of the T1 sample objects in the broad  $H\alpha$  flux ( $F_{bH\alpha}$ ) vs.  $\Delta v_{1000}$  plane. In each panel the coloured objects mark the 4% of the objects which are just above a given selection threshold. The top panel marks the positions of 4% of the objects where  $2.5 < \Delta F/\sigma_{\text{data}} < 6.4$ . The selection of  $\Delta F/\sigma_{\text{data}} > 2.5$  affects the lowest detectable  $\Delta v_{1000}$ . As  $F_{bH\alpha}$  decreases below  $10^{-14}$  erg s $^{-1}$  cm $^{-2}$ , the minimal detectable  $\Delta v_{1000}$  increases from 1 to 3. As the excess flux becomes lower, the line needs to be broader, otherwise it is lost in the glare of the NLR  $H\alpha$  component. Selection criterion (iii) in §2.5 of  $\Delta F'/\sigma_{\text{data}} > 4$  (not plotted), produces a similar effect.

The middle panel marks the positions of objects where  $2 < F_{\lambda, bH\alpha}/\epsilon_{\lambda} < 3.6$ . This selection produces a reverse effect. As  $F_{bH\alpha}$  decreases below  $10^{-14}$  erg s $^{-1}$  cm $^{-2}$ , the maximum detectable  $\Delta v_{1000}$  decreases from 10 to 3. As the excess flux density becomes lower, the line needs to be narrower, otherwise it is lost in the continuum noise.

The lower panel marks the positions of objects with  $10 < S/N < 17.1$ . Clearly, this selection criterion does not form any of the boundaries in the  $F_{bH\alpha}$  vs.  $\Delta v_{1000}$  plane. Of the 93 objects removed by eye (§2.5), 85% have  $F_{bH\alpha}$  and  $\Delta v_{1000}$  values which coincide with the  $2.5 < \Delta F/\sigma_{\text{data}} < 6.4$  objects (top panel), indicating the manual filtering does not introduce an additional boundary in this plane.

A similar analysis in the  $EW_{bH\alpha}$  vs.  $\Delta v_{1000}$  plane gives similar results. The minimum detectable  $\Delta v$  rises from 1 to 3 for  $EW_{bH\alpha} < 40\text{\AA}$ , and the maximum detectable  $\Delta v_{1000}$  decreases from 10 to 3 for  $EW_{bH\alpha} < 20\text{\AA}$ .

The objects in Fig. 2 extend towards  $\Delta v_{1000} = 1$ , in particular at  $F_{bH\alpha} < 10^{-13.5}$  erg s $^{-1}$  cm $^{-2}$ . Thus, the selection criterion  $\Delta v_{1000} > 1$  may exclude some type 1 AGN which have  $\Delta v_{1000} < 1$ . What is the possible missed fraction of AGN? We have excluded 149 ( $= 3579 \times 4\%$ ) objects that were fit with  $0 < \Delta v_{1000} < 1$ , due to possible confusion with emission from [N II], some of which may be true broad line AGN with  $\Delta v_{1000} < 1$ . The expected fraction of  $\Delta v_{1000} < 1$  AGN can be also estimated from the GH07 sample, where also 3% of the objects have  $\Delta v_{1000} < 1$ . Since a primary motivation of GH07 was to find broad line AGN with the lowest  $\Delta v_{1000}$  (low  $M_{BH}$ , see Greene & Ho 2004), it is likely to be as complete as possible, and implies the T1 sample likely misses the 3% of AGN in the SDSS with  $\Delta v_{1000} < 1$ , which may constitute part of the objects with an apparent  $\Delta v_{1000} < 1$  excluded here.

The  $\Delta v_{1000} < 25$  selection criterion has practically no effect

as very few objects reside at  $\Delta v_{1000} > 10$  (only one object with  $\Delta v_{1000} > 25$  is detected here<sup>3</sup>).

## 2.7 Additional surveys

We supplement the optical SDSS spectra in the T1 sample with photometric measurements in the UV, Near IR and X-ray. In order to include all AGN light, and to be consistent with the luminosity calibration of SDSS spectra (Adelman-McCarthy et al. 2008), we use luminosities from fixed apertures (different for each survey), corrected for the expected full-to-aperture light ratio of a point source. Note that for extended objects, the deduced luminosity is different than the luminosity obtained by a PSF fit to the image. This procedure is preferred over using galaxy-model fits, which perform poorly and differently for different AGN/host light ratios (e.g. §4.4.5 in Stoughton et al. 2002). Thus, comparing the AGN luminosity of our T1 objects to PSF luminosities of distant quasars (§3.3) is justified. Host luminosities are compared only with objects that were observed in the same way (§3.3 – §3.5). The specific details of each survey are described below, and the photometric luminosities of each object are listed in Table 1.

### 2.7.1 GALEX

The GALEX mission (Martin et al. 2005) imaged 2/3 of the sky with two broadband filters: far-UV (FUV; effective wavelength 1528Å) and near-UV (NUV; 2271Å). We use the GR6 data release, described in Morrissey et al. (2007) and in the MAST website<sup>4</sup>.

We search for GALEX detections in both bands, within 5'' of the T1 objects. In the case of multiple GALEX detections per object, we co-add the observations, weighted by exposure time. Figure 3 presents the detection fractions of the 89% of the T1 objects observed by GALEX, as a function of broad  $H\alpha$  flux ( $F_{bH\alpha}$ ) for point sources and for extended sources. The point sources in the sample are effectively complete in the UV. A trend of decreasing detection fraction with decreasing  $F_{bH\alpha}$  can be seen in the extended objects, but the detected fraction remains  $> 82\%$ . If we include NUV-only detections, 98% of the T1 objects observed by GALEX are detected.

We use the UV fluxes of the 6''-radius ‘aperture 4’, so that the residual correction for the PSF is  $< 10\%$  (Morrissey et al. 2007). We correct for the PSF and for Galactic reddening using  $A_{NUV}/E(B-V) = 8.2$  and  $A_{FUV}/E(B-V) = 8.24$  (Wyder et al. 2007). When a specific rest-frame UV wavelength is used, we interpolate (or extrapolate) a power-law from the measurements of the two bands.

### 2.7.2 2MASS

Near infrared photometric information was obtained from the point source catalog of the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006), which imaged 99.998% of the sky in the J (1.25  $\mu\text{m}$ ), H (1.65  $\mu\text{m}$ ), and K<sub>s</sub> (2.16  $\mu\text{m}$ ) bands to a sensitivity limit of 15 mag in the H-band (for point sources with  $S/N \sim 10$ ). We use the published luminosity from a 4''-radius aperture, fixed for the residual of the PSF.

For each object in the T1 sample, we search for a 2MASS counterpart using a matching radius of 2''. We filter 2MASS objects

<sup>3</sup> SDSS J094215.12+090015.8; see Wang et al. (2005).

<sup>4</sup> <http://galex.stsci.edu/GR6>

SDSS Name	$z$	$L_{\text{bH}\alpha}$	$\Delta v$	$K_s$	H	J	NUV	FUV	2 keV
J000202.95-103037.9	0.103	41.91	2320	44.13	44.19	44.23	43.83	43.83	42.58
J000338.94+160220.6	0.116	42.19	2080	44.14	44.23	44.16	43.59	43.49	42.49
J000410.80-104527.2	0.240	42.63	1370	44.55	44.63	44.47	44.65	44.69	43.22
J000611.55+145357.2	0.119	42.12	3300	44.18	44.24	44.30	43.94	44.03	42.60
J000614.36-010847.2	0.090	41.60	3960	43.77	43.86	43.89	43.30	43.26	0.

**Table 1.** The broad  $\text{H}\alpha$  luminosities (in  $\log \text{erg s}^{-1}$ ), widths (in  $\text{km s}^{-1}$ ) and 2MASS, GALEX and ROSAT  $\nu L_\nu$  luminosities (in  $\log \text{erg s}^{-1}$ ) of the T1 sample. GALEX and 2MASS luminosities are in observed frame. Non detections are denoted by ‘0’. For 2MASS and GALEX, a detection in all bands of the survey is required. An object that was not observed is denoted by ‘-1’ (GALEX only). The electronic version of the paper includes all 3 579 objects.

with a risk of nearby star contribution (‘cc\_flag’ = s,p; Cutri et al. 2003) or an upper limit in either of the three bands. We find that 96.5% of the T1 sample have such matches (Fig. 3). We correct for Galactic reddening using the  $A/E(B-V)$  values in Schlegel et al. (1998). When a specific rest-frame wavelength in the NIR is needed, a power-law is interpolated (or extrapolated) from the two nearest 2MASS bands.

### 2.7.3 ROSAT

The ROSAT All-Sky Survey (RASS; Voges et al. 1999, 2000) covers the entire celestial sphere in the 0.1 – 2.4 keV range with the Position Sensitive Proportional Counter (Pfeffermann et al. 1987) to a typical limiting sensitivity of  $\sim 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ . Typical positional uncertainties are  $10'' - 30''$ , and the detected source density is  $\sim 3 \text{ deg}^{-2}$ . Only 0.4% of the T1 sample was targeted for spectroscopy by the SDSS solely for being near a ROSAT source, and not as part of the complete SDSS Galaxy and Quasar surveys (see §2.1 and Stoughton et al. 2002). Therefore, we do not expect the T1 sample to be biased towards X-ray bright AGN.

We search for a RASS source within  $50''$  of each T1 object, expecting 7 objects to have false matches. We find matches to 1 545 (43%) of the T1 sample. The detection fraction vs.  $F_{\text{bH}\alpha}$  is shown in Figure 3. The strong trend of detection rate with  $F_{\text{bH}\alpha}$  implies that most of the objects without an X-ray detection are not a distinct class (‘X-ray weak’; Laor et al. 1997), but are rather AGN faint in all bands.

We use the PIMMS software to convert RASS count rates into X-ray fluxes, assuming a power-law X-ray spectrum with  $\alpha_x = 1.5$  (Laor et al. 1994, Schartel et al. 1996). Then, using the  $z$  of the optical match,  $\alpha_x = 1.5$ , and the Galactic  $N_{\text{H}}$  measurements of Stark et al. (1992), we derive the rest-frame  $\nu L_\nu(2 \text{ keV})$  of the matched objects, corrected for Galactic absorption.

## 3 RESULTS

### 3.1 The final sample

Figure 4 presents  $L_{\text{bH}\alpha}$  vs.  $z$  of the final T1 sample. The figure displays the 3 077 objects at  $z < 0.2$ . Empty circles mark the 423 objects (14%) that also appear in the QCV sample. The QCV sample objects mostly reside at  $\log L_{\text{bH}\alpha} \approx 42.5 - 43$ , while the T1 sample extends further down to  $\log L_{\text{bH}\alpha} \approx 40$ . The T1 sample includes additional 502 objects at  $0.2 < z < 0.31$  not displayed in Fig. 4. In this range there is an overlap of 88% of our objects with the QCV sample, but our sample includes only 24% of the QCV objects in this range. This small fraction is due to our requirement of high quality data near  $\text{H}\alpha$ , which is satisfied by a decreasing fraction of the objects as  $\text{H}\alpha$  shifts into the noisy red end of the SDSS spectra (§2.1).

The total T1 sample extends over the range  $\log L_{\text{bH}\alpha} \sim 40 - 44$ , and becomes distinct from the QCV sample at  $z < 0.2$ .

Figure 5 compares the ranges of  $L_{\text{bH}\alpha}$  spanned by the T1 sample with the  $z < 0.35$  QCV sample (where  $\text{H}\alpha$  is observable). Our sample extends down by nearly two orders in magnitude in  $L_{\text{bH}\alpha}$ , compared to the QCV sample. Specifically, the T1 sample includes 1 878 objects at  $L_{\text{bH}\alpha} < 10^{42} \text{ erg s}^{-1}$ , compared to 207 in QCV, and 297 objects at  $L_{\text{bH}\alpha} < 10^{41} \text{ erg s}^{-1}$ , compared to 5 in QCV. We also show the Ho et al. (1997a) sample (44 objects, updated values in Ho 2003), which extends below  $10^{38.5} \text{ erg s}^{-1}$ , while the T1 sample terminates just below  $10^{40} \text{ erg s}^{-1}$ .

In this paper we explore the emission properties of the T1 sample as a function of various parameters, including  $M_{\text{BH}}$  and  $L/L_{\text{Edd}}$ . We derive these parameters from  $\text{H}\alpha$ , following Greene & Ho (2005) and Kaspi et al. (2005). We use the relation between the BLR radius and UV luminosity in Kaspi et al. (2005),

$$R_{\text{bH}\beta} = (17.6 \pm 0.26) \times \nu L_\nu(1450\text{\AA})_{44}^{0.545 \pm 0.06} \text{ light days}, \quad (1)$$

We then use the relation  $\nu L_\nu(1450\text{\AA}) = 33 \times L_{\text{bH}\alpha}^{1.0}$  (eq. 5 below), and Greene & Ho (2005, eq. 3 there) to convert  $\text{FWHM}(\text{broad H}\beta)$  to  $\text{FWHM}(\text{broad H}\alpha)$ , which gives

$$m \equiv \log \frac{M_{\text{BH}}}{M_\odot} = 7.4 + 2.06 \log \Delta v_{1000} + 0.545 \log L_{\text{bH}\alpha,44}. \quad (2)$$

We use  $L_{\text{bH}\alpha}$  to obtain  $L_{\text{bol}}$  using the relation  $L_{\text{bol}} = 130 \times L_{\text{bH}\alpha}^{1.0}$  (eq. 6 below), which gives

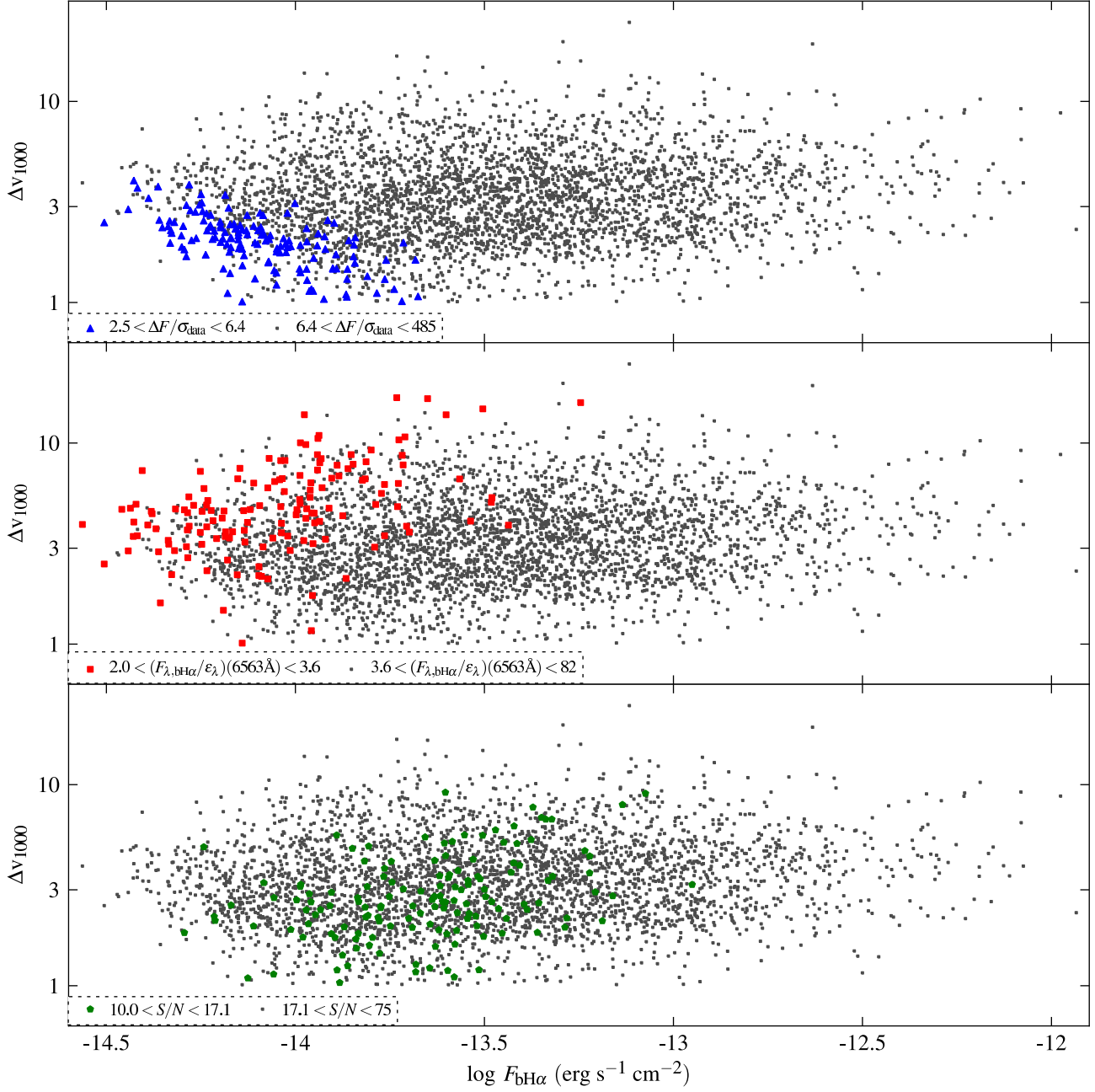
$$l \equiv \log \frac{L}{L_{\text{Edd}}} = 0.6 - 2.06 \log \Delta v_{1000} + 0.455 \log L_{\text{bH}\alpha,44} \quad (3)$$

Note that eq. 2 is equivalent to eq. 6 in Greene & Ho (2005), but is slightly different as it uses the Kaspi et al. (2005) relation, and the relations between  $L_{\text{bH}\alpha}$  and continuum luminosity obtained below.

It is important to understand how the sample selection criteria affect the observed range of various AGN parameters. Figure 6 shows the observed distribution of the T1 sample objects in the  $L_{\text{bH}\alpha}$  vs.  $\Delta v$  plane. In the top panel, the larger (coloured) dots in this figure present the objects close to our selection cutoffs (see Fig. 2). The  $m$  and  $l$  diagonal lines are derived from eqs. 2 and 3. It is clear that  $m \lesssim 6$  objects are heavily selected against due to the  $\Delta F/\sigma_{\text{data}} > 2.5$  criterion, while the  $F_{\lambda, \text{bH}\alpha}/\epsilon_\lambda > 2$  criterion selects against  $l \lesssim -2.8$  objects.

Also, we can now identify the effect of the rarity of active very massive black holes ( $m \gtrsim 9$ ) in the local universe, which manifests as a lack of high- $\Delta v$ , high- $L_{\text{bH}\alpha}$  objects. Another limit on the distribution is due to the Eddington luminosity, which leads to an increasing minimum  $\Delta v$  with  $L_{\text{bH}\alpha}$  for  $L_{\text{bH}\alpha} > 10^{42.5} \text{ erg s}^{-1}$ .

The lower panel shows  $\Delta v$  as a function of  $L_{\text{bol}}$ . Here we add the 8 185 QCIV objects from Shen et al. (2008), where the broad  $\text{H}\beta$  is available ( $z < 0.66$ ). We fix for a mean offset of 0.13 dec (with a dispersion of 0.10 dec) between the Shen et al.  $\text{FWHM}(\text{H}\beta)$

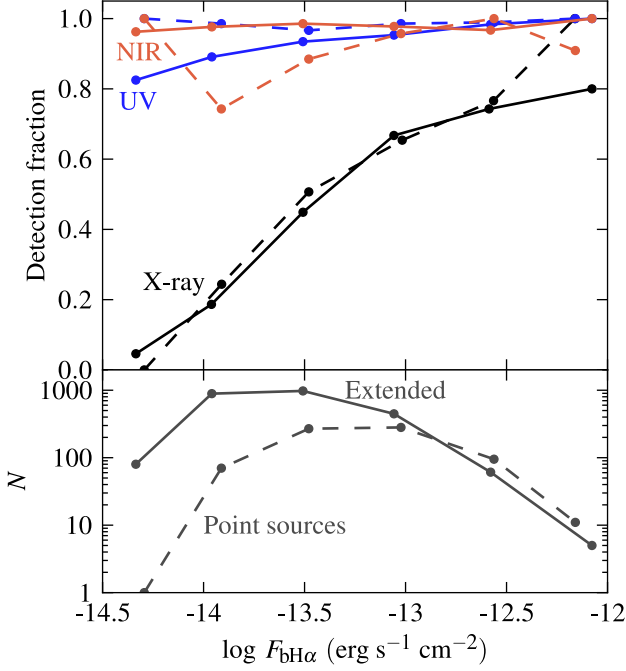


**Figure 2.** The selection criteria effects on the distribution of the T1 objects (small gray dots) in the broad  $H\alpha$  flux vs. FWHM plane. Large coloured markers are used to plot the 4% of the T1 objects which are just above the following three selection thresholds. (**Upper panel**) The  $\Delta F/\sigma_{\text{data}} > 2.5$  selection criterion. At  $F_{bH\alpha} < 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ , this selection increases the minimal detectable  $\Delta v_{1000}$  from 1 to 3. (**Middle panel**) The  $F_{\lambda,bH\alpha}/\epsilon_{\lambda} > 2$  selection criterion. At  $F_{bH\alpha} < 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ , this selection decreases the maximal detectable  $\Delta v_{1000}$  from 10 to 3. The (**Lower panel**)  $S/N > 10$  selection criterion. This selection criterion does not form any boundary in the  $F_{bH\alpha}$  vs.  $\Delta v_{1000}$  plane. The T1 objects extend down to  $\Delta v_{1000} = 1$ , in particular at  $F_{bH\alpha} < 10^{-13.5} \text{ erg s}^{-1} \text{ cm}^{-2}$ , and objects with  $\Delta v_{1000} < 1$  are excluded ( $\sim 4\%$ , see text). The  $\Delta v_{1000} < 25$  criterion has practically no effect, as objects with  $\Delta v_{1000} > 15$  are extremely rare. Our  $F_{bH\alpha}$  detection limit is lowest for  $\Delta v_{1000} \sim 3$  objects.

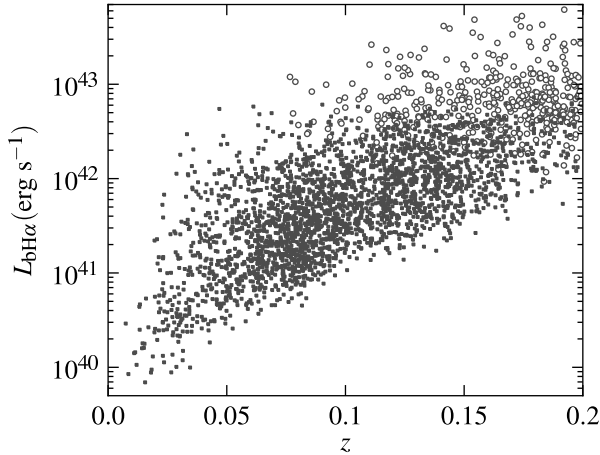
measurement and our  $\text{FWHM}(H\alpha)$  measurement, derived based on common objects. The value of  $L_{\text{bol}}$  in the T1 sample is derived from eq. 6, and in the Shen et al. sample  $L_{\text{bol}}$  is derived from  $\nu L_{\nu}(5100\text{\AA})$ . The plots show clearly how the Eddington luminosity limit extends to quasar luminosities. The Shen et al.  $\text{FWHM}$  measurements extend to higher luminosities using  $\text{FWHM}(\text{Mg II})$  or  $\text{FWHM}(\text{C IV})$  available for higher  $z$  objects. We avoid utiliz-

ing these measurements since the relation of  $\text{FWHM}(\text{Mg II})$  to  $\text{FWHM}(H\beta)$  is not linear, and the  $\text{FWHM}(\text{C IV})$  shows only a very weak correlation with  $\text{FWHM}(\text{Mg II})$  (e.g. Shen et al. 2008, Fine et. al. 2010). Note that with increasing  $L_{\text{bol}}$  the range in the observed  $\Delta v$  shrinks due to the upper limits on  $l$  and  $m$ . A similar effect is seen for  $\text{Mg II}$  (Fine et. al. 2008), and  $\text{C IV}$  (Fine et. al. 2010). In fact, no broad line objects are expected to exist at a high

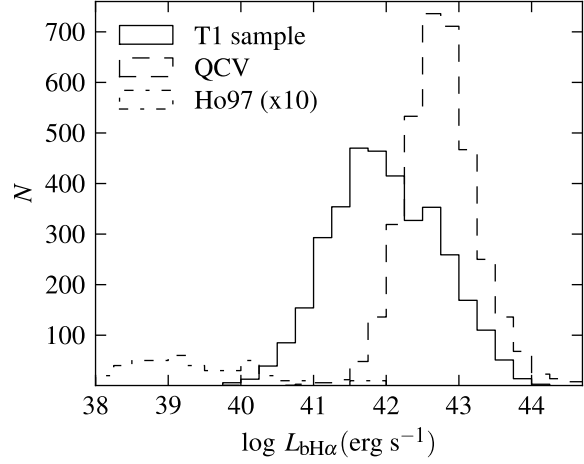




**Figure 3.** The detection fractions of the T1 objects in the GALEX (UV, blue), 2MASS (NIR, red) and ROSAT (X-ray, black) surveys, as a function of  $F_{bH\alpha}$ , for extended and point like sources. Detections in all bands of each survey is required. Solid/dashed lines represent extended/point source objects. The lower panel shows the number of T1 point-like and extended objects per 0.5 decade  $F_{bH\alpha}$  bin. The GALEX detection fraction is relative to the 89% of T1 objects observed. Note the nearly complete 2MASS and GALEX detections. When analyzing the X-ray luminosities (§3.6 and §3.9), we utilize only  $F_{bH\alpha} > 10^{-13.5} \text{ erg s}^{-1} \text{ cm}^{-2}$  objects, where the detection fraction is  $> 50\%$ . Note that most of the lowest  $F_{bH\alpha}$  T1 AGN reside in extended objects. Their bright host enables them to pass the SDSS flux limit, despite their low  $F_{bH\alpha}$ .



**Figure 4.** The  $L_{bH\alpha}$  versus  $z$  distribution for the T1 sample at  $z < 0.2$  (3 077 objects). Empty circles mark the 423 objects (14%) that also appear in the SDSS quasar catalog (QCV). Note that the QCV sample terminates at  $L_{bH\alpha}$  of a few  $10^{42} \text{ erg s}^{-1}$ , while the T1 sample extends further down to  $10^{40} \text{ erg s}^{-1}$ .



**Figure 5.** Comparison of  $L_{bH\alpha}$  ranges of different broad line AGN samples. We include only  $z < 0.35$  objects of the QCV sample, where  $H\alpha$  is observable. Although the T1 sample extends down by two orders of magnitude in  $L_{bH\alpha}$ , compared to QCV, it does not extend as low as the Ho et al. (1997a) sample, which reaches  $L_{bH\alpha} < 10^{38.5} \text{ erg s}^{-1}$  (magnified ten-fold for clarity).

enough  $L_{bol}$  due to the Eddington limit. This merely reflects the fact that the minimal required  $M_{BH}$  increases with  $L_{bol}$ , and AGN above some  $M_{BH}$  just do not exist.

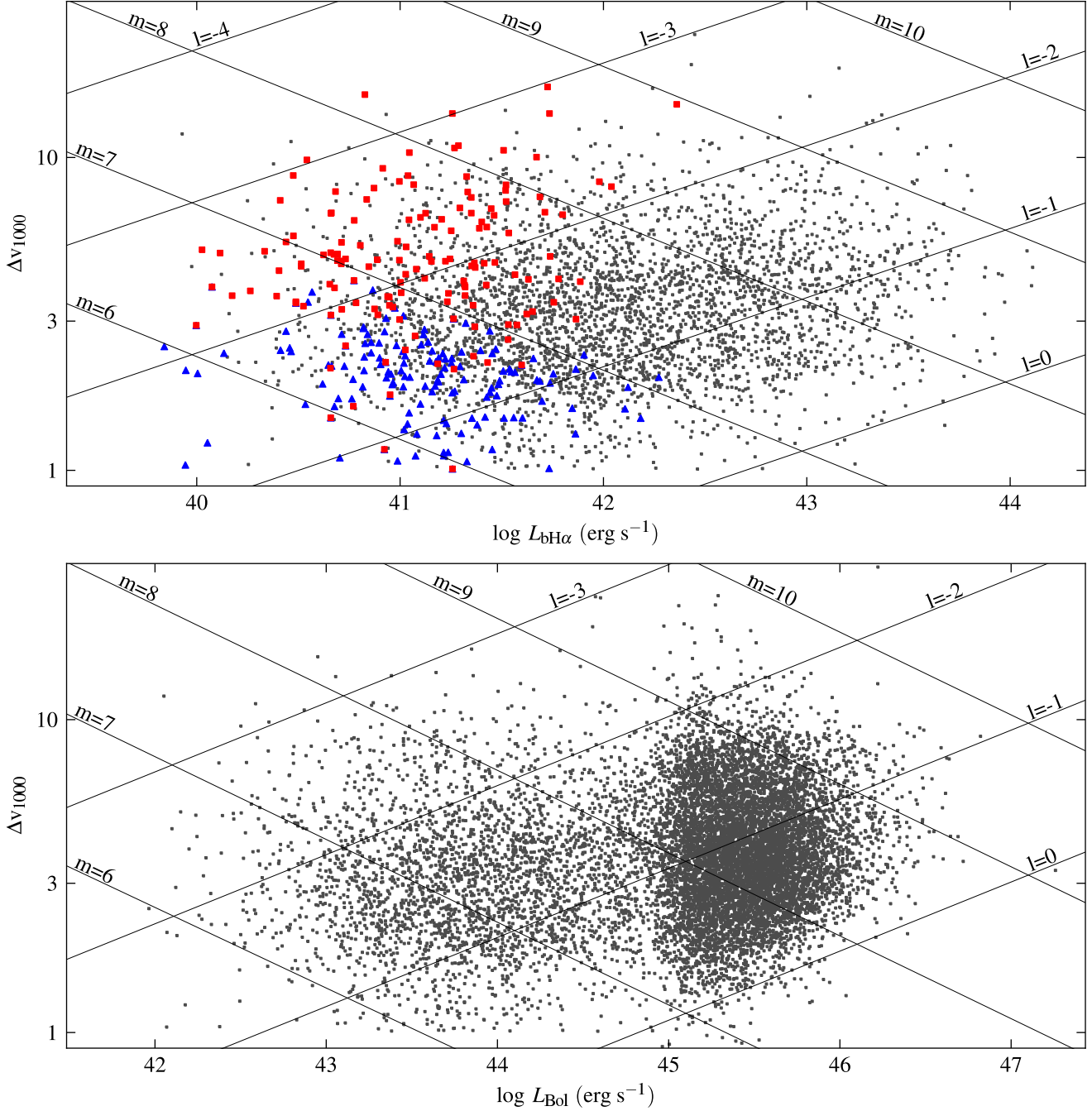
### 3.1.1 The $\Delta v$ distribution

How broad do the broad lines get? The lower panel of Fig. 6 shows there is a steep decline in the number of objects with  $\Delta v > 10000 \text{ km s}^{-1}$ . The decline is independent of  $L_{bol}$ , over four orders of magnitude in  $L_{bol}$ . The decline does not result from our selection criteria, as shown in the upper panel. Figure 7 displays histograms of the FWHM distribution  $dN/d\log \Delta v$  vs.  $\Delta v$ , at four luminosity bins. The distributions are remarkably similar, showing a roughly linear decline of  $\log(dN/d\log \Delta v)$  vs.  $\Delta v$ , or equivalently  $dN/d\log \Delta v \propto e^{-\Delta v/\Delta v_0}$ , with  $\Delta v_0 \approx 2700 \text{ km s}^{-1}$ . The distributions peak at  $\Delta v \approx 2500 - 4000 \text{ km s}^{-1}$ , and drops steeply for smaller values. The sharp drop at the smallest  $\Delta v$ , and the increase in  $\Delta v$  of the peak position with  $L_{bol}$ , is due to the Eddington limit ( $l = 0$ , see Fig. 6), which gives a minimal  $\Delta v \propto L_{bH\alpha}^{0.22}$  (eq. 3). The similar distributions for  $\Delta v$  above the peak is surprising, as the  $\Delta v$  distributions are set by the distribution of the number of objects vs.  $l$  and vs.  $m$ . It is not clear why these two distributions combine to yield the same  $\Delta v$  distributions at different  $L_{bol}$ . It may imply that the distributions of  $\Delta v$  is directly set by the value of  $\Delta v$ , but it is not clear what may be the physical mechanism behind a dependence on  $\Delta v$  only (see further discussion in §4.1).

### 3.1.2 The $L/L_{Edd}$ distribution

Figure 8 compares the Eddington ratio distributions of the T1 sample with the BQS (using the BG92  $z < 0.5$  subsample, with values taken from Baskin & Laor 2005) and the Kollmeier et al. (2006, hereafter K06) quasar samples. Both the BQS and K06 were selected as point sources, in contrast with the T1 sample, which includes both point sources and extended sources. The T1 sample is centred around  $l \approx -1.3$ , while the BQS and the K06 samples are centred around  $l \approx -0.6$ . To determine if the higher  $l$  distributions



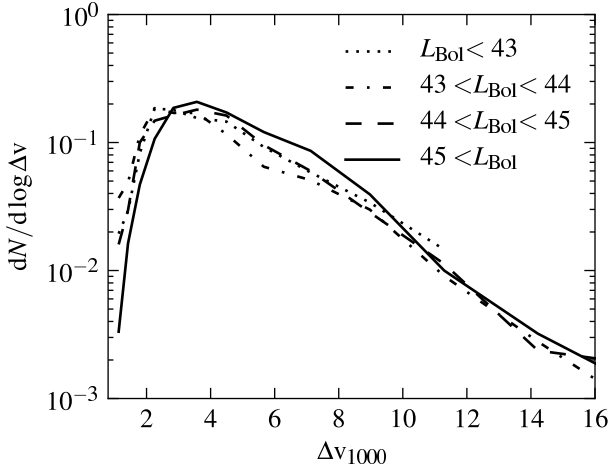


**Figure 6. (Upper panel)** The distribution of the T1 sample objects in the  $L_{\text{bH}\alpha}$  vs.  $\Delta v$  plane. Large coloured markers are as in Fig. 2. Objects with  $m \lesssim 6$  or  $l \lesssim -2.8$  are selected against. The Eddington limit ( $l = 0$ ) sets a minimum  $\Delta v$  with increasing  $L_{\text{bH}\alpha}$  for  $L_{\text{bH}\alpha} > 10^{42.5}$  erg s<sup>-1</sup>. This, together with the rarity of AGN with  $m > 9$ , leads to a decrease in the range of observed  $\Delta v$  values with increasing  $L_{\text{bH}\alpha}$ . **(Lower panel)** The distribution of the T1 sample + QCIV sample (values from Shen et al. 2008, see text) in the  $L_{\text{bol}}$  vs.  $\Delta v$  plane ( $L_{\text{bol}}$  is from eq. 6). The Eddington limit is clearly seen. Note there is a similar steep decline in the number of objects with  $\Delta v > 10000$  km s<sup>-1</sup>, at all luminosities. The decline is not set by our selection criteria.

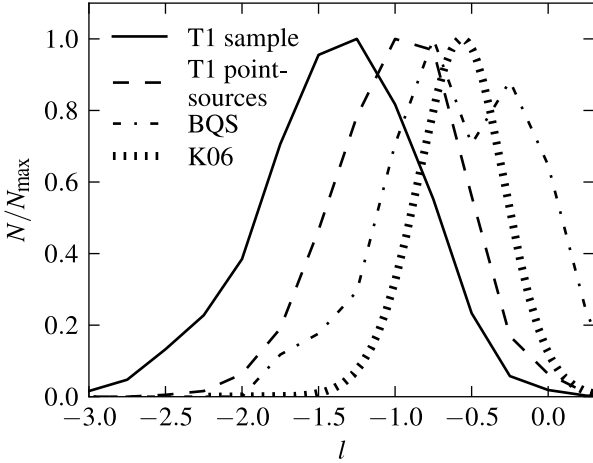
are driven by the point source selection, we show the  $l$  distribution of the point sources in our T1 sample. The T1 point sources are also shifted to higher  $l$  values. This clearly demonstrates that the higher  $l$  values of BQS and K06 samples are largely driven by the point-like selection effect. This trend is well understood by the  $M_{\text{BH}}$  versus bulge luminosity relation (Magorrian et al. 1998), which implies that  $L/L_{\text{Edd}}$  is equivalent to  $L_{\text{AGN}}/L_{\text{bulge}}$ . Thus quasars, i.e. objects

which are selected by their quasi-stellar appearance, are by their selection high  $L/L_{\text{Edd}}$  AGN.

Below we compare T1 hosts to other galaxy types (in §3.1.4 and §3.5), where we use the subsample of 2 766 T1 AGN which are part of the SDSS main galaxy sample (Strauss et al. 2002). These objects are selected to have extended morphologies, and therefore by selection exclude high  $L/L_{\text{Edd}}$  objects (Fig. 8).



**Figure 7.** The FWHM distribution of the combined T1 + QCIV sample (Shen et al. 2008), presented in Fig. 6, at different  $L_{\text{bol}}$  bins. The distributions are similar. The origin of this similarity is not clear. The  $\Delta v$  distributions should be set by the distribution of  $m$  and  $l$  values. Either these distributions somehow lead to a  $\Delta v$  distribution which is independent of  $L_{\text{bol}}$ , or it may imply that  $\Delta v$ , rather than  $m$  and  $l$ , sets the observed  $\Delta v$  distribution through some unknown mechanism. Note the effect of the Eddington limit at  $\Delta v_{1000} < 3$ , which increases the minimal  $\Delta v$  and the peak position with increasing  $L_{\text{bol}}$ .



**Figure 8.** A comparison of the Eddington ratio distribution in various samples. The  $l$  bin size is 0.25 decades. Both the BQS and the K06 samples were selected to be point sources. The T1 distribution is shifted to significantly lower  $l$  values. To test the origin of this shift we plot the distribution of the T1 objects originally identified as point like sources, which is clearly shifted to higher  $l$  values. Therefore, the notion that AGN tend to shine at  $l \sim -0.5$  (K06) reflects the point like selection criterion. AGN termed quasi-stellar have a high  $l$  by construction.

### 3.1.3 Other low luminosity type 1 samples

Below we compare the completeness of the T1 sample to the other SDSS based samples of low luminosity type 1 AGN of Hao et al. (2005a), Vanden Berk et al. (2006) and GH07.

The Vanden Berk et al. (2006) sample is based on the SDSS DR3 and includes 4666 objects. The AGN selection criterion is the detection of at least one emission line profile with FWHM  $> 1000 \text{ km s}^{-1}$ . This criterion excludes AGN where the narrow

component dominates the Balmer line profiles (e.g. panels ‘b’ and ‘c’ of Fig. 1). As a result, the Vanden Berk et al. (2006) sample misses 27% of the T1 objects present in SDSS DR3.

The Hao et al. (2005a) sample is based on SDSS DR2 and includes 1317 objects, of which 52.5% (692 objects) entered our parent sample (passing our  $S/N > 10$  and bad pixel criteria). Hao et al. fit the narrow and broad  $H\alpha$  profile with a simpler procedure than described in §2, using one or two Gaussians. Of their objects that entered our parent sample, 27% were rejected due to a lack of broad line flux (§2.3), and further 20% were rejected based on the excess flux shape (§2.5). A similar percentage of 47% of the Hao et al. (2005a) sample were rejected by GH07. We therefore suspect that these broad  $H\alpha$  detections by Hao et al. are not significant.

The GH07 sample is based on SDSS DR4 and includes 8495 objects. Our selection method of excess broad flux near  $H\alpha$  is similar to the GH07 selection criterion. The broad  $H\alpha$  detection rates in the T1 and GH07 samples are both 1.5%. However, GH07 fit the broad  $H\alpha$  feature in a smaller velocity range of  $-7.4 < v_{1000} < 6.3$  from the  $H\alpha$  peak position, compared to  $-14.3 < v_{1000} < 14.5$  used in the T1 sample. Thus, we find that 1.6% of our objects (58/3596) have  $\Delta v_{1000} > 10$ , compared to 0.51% (43/8495) in GH07. Also, 17% of the 2273 GH07 which appear in our parent sample, were rejected from the T1 sample, potentially due to more strict selection criteria applied here. Comparing the measured  $\Delta v$  values in individual objects which appear in both samples, we find that the T1 values are on average a few percent larger, with a mean offset of  $\Delta v - \Delta v_{\text{GH07}} = 222 \text{ km s}^{-1}$ . The RMS dispersion of the differences in the individual  $\Delta v$  values is  $875 \text{ km s}^{-1}$ , and the RMS of the fractional difference in  $\Delta v$  is 20% (0.13 for  $\log L_{\text{bH}\alpha}$ ). As noted in §2.6, 3% of GH07 have  $\Delta v_{1000} < 1$ , while such objects are excluded from the T1 sample.

### 3.1.4 The fraction of SDSS objects which are T1

It is also interesting to compare the T1 sample to samples of type 2 AGN and samples of inactive galaxies. We use samples derived from the SDSS, to avoid biases due to different selection effects of other surveys. We use the classifications of Brinchmann et al. (2004, hereafter B04), who classified SDSS galaxy spectra according to the equivalent width and line ratios of narrow emission lines. Galaxies with narrow lines were divided according to their positions in the BPT diagrams (Baldwin, Phillips & Terlevich 1981; Veilleux & Osterbrock 1987). This classification was performed on all DR7 spectra in which the galaxy is dominant<sup>5</sup>. Objects which are clearly type 1 AGN were excluded, to avoid AGN contamination of the host galaxy colours. So, the B04 classification includes only type 2 AGN. However, as mentioned in Kauffmann et al. (2003a), objects with weak broad lines were not excluded. Since our T1 sample reaches low  $L_{\text{AGN}}/L_{\text{host}}$  values, it can include apparently type 2 AGN from B04. We find that out of 23279 objects classified as ‘AGN’ by B04, 12340 appear in our parent sample (following the  $S/N > 10$  and bad pixel cuts, §2.1). Of these, 472 (4%) are part of our T1 sample. This fraction is comparable with the Kauffmann et al. (2003a) estimate of 8% weak type 1 objects in their type 2 SDSS DR1 sample, which were not specifically identified.

It is interesting to explore the fraction of T1 objects classified by B04 as non AGN. In addition to the 472 T1 objects classified as type 2 AGN, 261 T1 objects were classified as composites, 42

<sup>5</sup> Results available at <http://www.mpa-garching.mpg.de/SDSS/DR7/>

as star forming galaxies (SFG), 20 as LINERs, and one as a non-emission line galaxies (NEG). So, a total of 324 objects out of 796, or 41%, of low luminosity broad line AGN are not clearly identified as AGN based on their narrow lines. In a following paper we measure the narrow lines for all the T1 objects, and discuss their BPT based identifications.

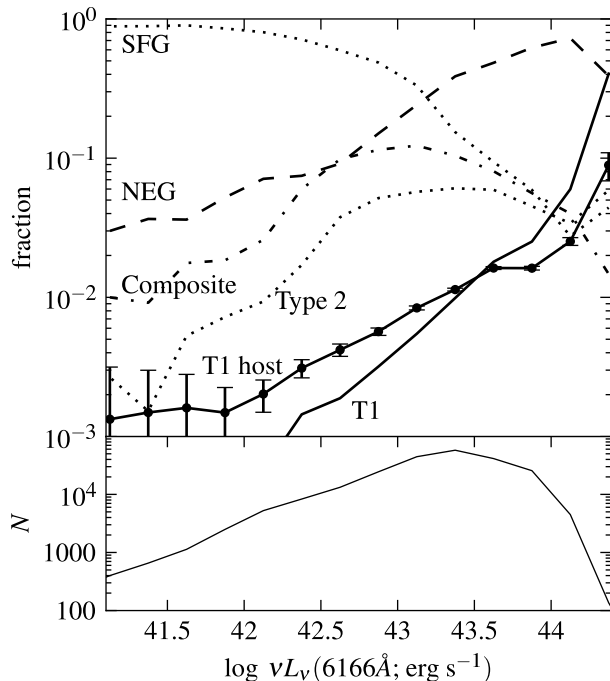
Figure 9 presents the classification of SDSS galaxies as a function of the spectroscopically measured  $\nu L_V(6166\text{\AA})$ , which corresponds to the SDSS  $r$  band (but measured within a  $3''$ -diameter fibre). We use all objects in the parent sample (§2.1) which are part of the SDSS main galaxy sample (Strauss et al. 2002). The T1 classification includes 2766 objects (about 0.1% of the objects were not classified neither here nor by B04, and are disregarded). For the T1 objects we derive  $\nu L_V(6166\text{\AA})$  from the mean of 20-pixels around  $6166\text{\AA}$ . For non-T1 objects we convert the synthetic<sup>6</sup>  $r$ -magnitudes published by B04 to  $\nu L_V$ . We correct the B04 measurement by  $-0.05$  dec, which is the mean offset between the two methods in common objects.

The classification fractions in Fig. 9 are somewhat different from B04 since some of the classifications depend on the emission lines S/N, which are higher on average in the  $S/N > 10$  parent sample presented here. Note that SFG heavily dominate ( $> 90\%$ ) at the lowest luminosities, and NEG dominate at the highest luminosities. The fraction of composite galaxies and type 2 AGN peaks at intermediate luminosities. In the T1 sample the luminosity can be dominated by the AGN. We therefore remove the AGN contribution to  $\nu L_V(r)$  using the result presented in §3.4 (eq. 5, 6% of T1 objects with an implied negative host are disregarded). The fraction of AGN-subtracted T1 objects is marked as ‘T1 host’. The error bars denote the error in the fraction due to the dispersion of eq. 5.

Remarkably, the T1 fraction follows well the NEG fraction, or equivalently, their galaxy luminosity distribution is similar. This suggests the two populations are related. If NEG are the host galaxies of type 1 AGN (with extended morphologies), then a fixed fraction of  $\sim 3\%$  NEG host broad line AGN, at the level detectable in this study. Alternatively, if the T1 occur in SFG, then the fraction of SFG which host broad line AGN increases sharply from  $2 \times 10^{-3}$  to about unity with luminosity. Further results on the T1 hosts are given in §3.5, and discussed in §4.3.

### 3.2 The IR to X-ray luminosity versus $L_{\text{bH}\alpha}$

Below we explore the correlation of the continuum luminosity in the X-ray, UV, optical, and near IR with  $L_{\text{bH}\alpha}$ , which provides a measure of the AGN luminosity without host contribution. We divide the T1 sample into  $L_{\text{bH}\alpha}$  bins, each half a decade wide (the highest bin,  $L_{\text{bH}\alpha} > 10^{43.8} \text{ erg s}^{-1}$ , includes only 14 objects and is ignored). Table 2 lists the number of objects per bin and the fraction of objects with detections in the various bands. The NIR detection is nearly complete, with a mean detection fraction of 96%. The mean UV detection is high (93%), remaining  $\geq 80\%$  even at the lowest  $L_{\text{bH}\alpha}$  bins. The high detection rates ensures the derived mean continuum properties are not significantly biased in the UV and in the NIR. The mean X-ray detection fraction is 43%, and drops below 50% for  $L_{\text{bH}\alpha} < 42.3 \text{ erg s}^{-1}$ . The effects of the X-ray band detection rate are discussed in §3.6. A search for matched radio detections in the FIRST survey revealed an average detection rate of only 18%. We therefore preferred not to explore further relations with the radio properties.



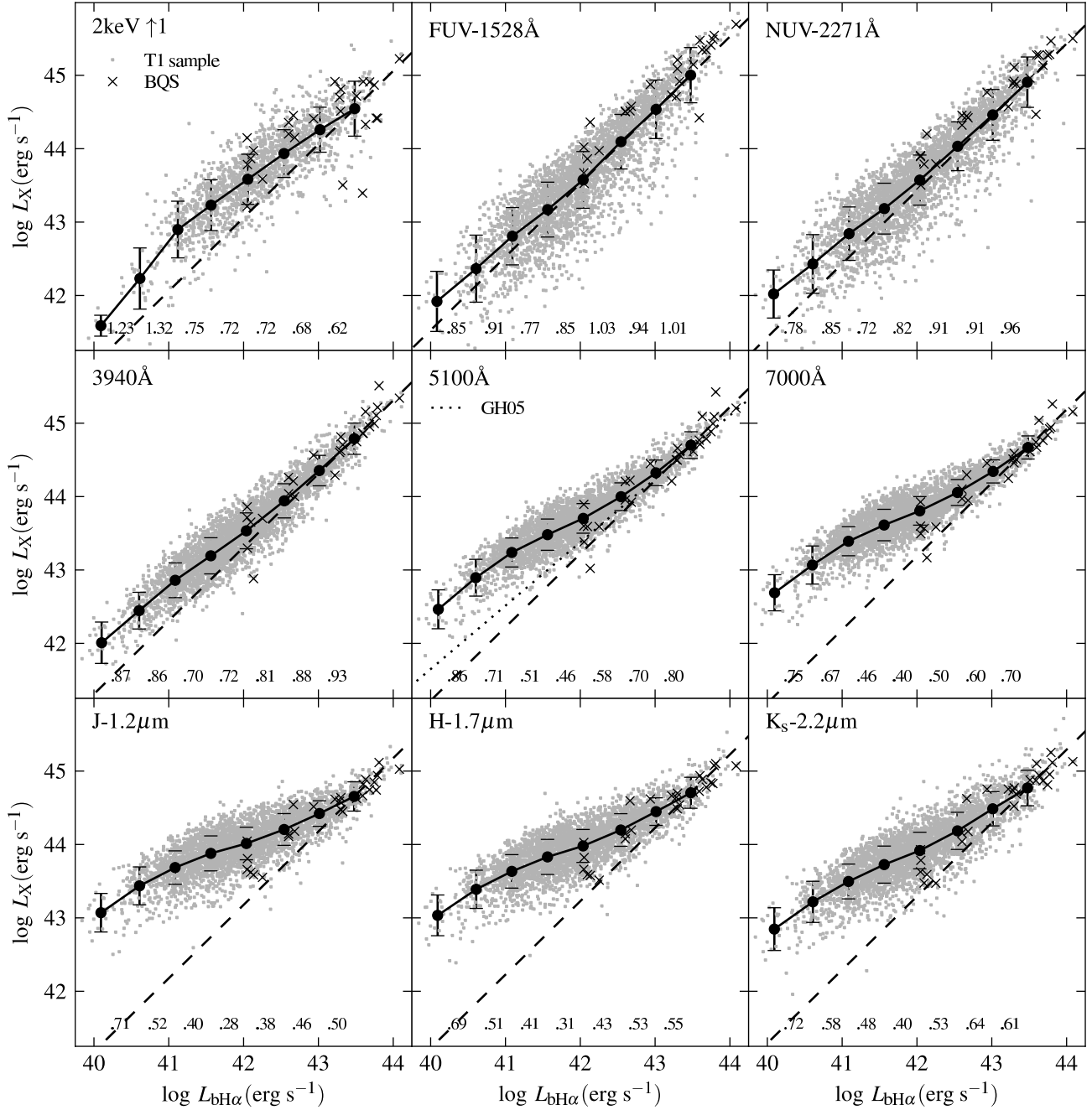
**Figure 9.** The distribution of spectral types of 230 413  $z < 0.2$  galaxies in the parent sample (§2.1), as a function of the spectroscopy derived luminosity (roughly at the  $r$ -band). The type 2, SFG, NEG and composite classifications (based on the EW of the narrow lines and the BPT diagrams) are taken from B04. At low luminosities  $> 90\%$  of the SDSS galaxies are SFG, while at high luminosities the NEG dominate. The T1 curve is based on the 2766 extended objects from the T1 sample. The ‘T1 host’ band is corrected for the AGN contribution to the luminosity. The error bars are derived from the dispersion in eq. 5. Note that the T1 host luminosity constitute a nearly fixed fraction of  $\sim 3\%$  of the NEG luminosity distribution. This may suggest that the T1 reside in NEG. The lower panel plots the total number of objects per  $\nu L_V$  bin of size 0.25 dec.

Figure 10 shows the correlation of  $L_{\text{bH}\alpha}$  vs.  $\nu L_V$  in nine rest frame bands from the X-ray to the NIR. The 2 keV luminosity is from ROSAT, 1528Å and 2271Å from GALEX, 3940Å, 5100Å, and 7000Å from SDSS, and 1.2μm, 1.7μm, and 2.2μm (J, H, K<sub>s</sub> bands) from 2MASS. The optical  $\nu L_V$  are measured on 20-pixel windows from the SDSS spectrum. The longest rest-frame wavelength is chosen so that it is available in all objects<sup>7</sup>. The non-optical bands are also in rest wavelengths, calculated from the photometric values as described in §2.7. The average and the dispersion in  $\log \nu L_V$  in each  $L_{\text{bH}\alpha}$  bin are marked in the plot. The ROSAT upper limits are shown in §3.6. The upper limits for the small fraction of objects with no GALEX detections are not plotted (it is not obvious how to derive them from the exposure time). The fraction of upper limits for 2MASS is tiny, so we do not plot the upper limits.

Fig. 10 shows there is a larger dispersion in the UV and X-ray luminosities for a given  $L_{\text{bH}\alpha}$ , compared to the dispersion in the optical and NIR bands. This likely reflects the larger variability in these bands. The relations of  $\log \nu L_V$  vs.  $\log L_{\text{bH}\alpha}$  in the different bands clearly do not follow a simple linear (i.e. single power-law) relation. Instead of attempting to fit the distributions with a more complicated relation, say a double power-law, we

<sup>6</sup> i.e., the spectrum convolved with the filter pass function.

<sup>7</sup> Objects with  $< 6$  good pixels in the relevant window are excluded (0.6%, 2% and 4% in 3940Å, 5100Å, and 7000Å, respectively).



**Figure 10.** The correlation of  $L_{bH\alpha}$  vs. observed luminosity from the rest frame X-ray to the NIR (the y-axis values in the X-ray panel are shifted upward by 1 decade for presentation purpose). The black dots and the associated error bars mark the mean and the dispersion of  $\nu L_\nu$  in each 0.5 decade wide  $L_{bH\alpha}$ -bin. The dispersions in the UV and X-ray bands are larger than in the optical and NIR bands, probably reflecting the higher variability. The solid lines connect the averages of adjacent bins, with the local slopes written underneath. Dashed lines show linear relations, normalized by the highest  $L_{bH\alpha}$ -bin. In the luminous bins in the FUV, NUV and 3940Å panels, the slopes are consistent with a linear relation, in all other bands and luminosities the slopes are flatter (except the X-ray panel at the lowest  $L_{bH\alpha}$  bins). There is a transition in all bands, excluding X-ray, from a steeper slope to a flatter slope with decreasing  $L_{bH\alpha}$ . The transition  $L_{bH\alpha}$  increases with wavelength, up to a maximum of  $\approx 10^{43}$  erg s $^{-1}$  at 1.2 $\mu$ m. As shown below (Fig. 11), the transition occurs due to the host contribution, which peaks at 1.2 $\mu$ m. The 5100Å panel also presents the  $L_{bH\alpha} - L_{AGN}(5100\text{\AA})$  relation from Greene & Ho (2005), derived for a sample selected to minimize the host contribution. Objects that appear in the BQS sample are marked by a cross. Note that for  $\lambda \geq 5100\text{\AA}$  the BQS objects are near the linear relation, consistent with their point like selection criterion which excludes significant host contribution. At shorter  $\lambda$  the BQS are above the line, consistent with their blue optical colour selection. Note that the EW of H $\alpha$ , as implied by the 7000Å panel, shows an apparent ‘inverse Baldwin effect’ if the host light is not properly subtracted.

$L_{\text{bH}\alpha}$ range (log erg s $^{-1}$ )	mean $L_{\text{bH}\alpha}$ (log erg s $^{-1}$ )	$N$	mean $z$	resolved fraction	detection fractions			R06 scaling (dec)
43.3 - 43.8	43.48	136	0.237	0.03	0.99	1.00	0.69	-0.75
42.8 - 43.3	43.01	406	0.205	0.33	0.94	0.98	0.63	-1.22
42.3 - 42.8	42.54	662	0.163	0.65	0.95	0.96	0.56	-1.68
41.8 - 42.3	42.04	859	0.119	0.86	0.97	0.94	0.47	-2.19
41.3 - 41.8	41.56	853	0.094	0.96	0.97	0.92	0.33	-2.67
40.8 - 41.3	41.09	488	0.070	0.99	0.99	0.90	0.21	-3.14
40.3 - 40.8	40.60	138	0.041	1.00	0.96	0.89	0.20	-3.62
39.8 - 40.3	40.10	23	0.021	0.96	0.91	0.80	0.22	-4.12

**Table 2.** Detection fractions in the eight  $L_{\text{bH}\alpha}$  bins. The morphology is taken from the SDSS (Stoughton et al. 2002). The detection fractions are in the 2MASS (NIR), GALEX (UV) and ROSAT (X-ray) surveys, and require a detection in all bands of the survey. The R06 scaling is discussed in §3.4.

just form a ‘ridge’ line by connecting the averages of the adjacent  $L_{\text{bH}\alpha}$  bins. The slope in each segment (noted in the bottom of each panel) is flatter at lower  $L_{\text{bH}\alpha}$ , excluding the X-ray panel (further discussed in §3.6). The change of slope in the FUV occurs at  $L_{\text{bH}\alpha} \simeq 10^{41.5}$  erg s $^{-1}$ . The  $L_{\text{bH}\alpha}$  where the slope change occurs increases at longer wavelengths, and reaches a maximum of  $\simeq 10^{43}$  erg s $^{-1}$  at  $1.2\mu\text{m}$ . It then decreases back to  $\simeq 10^{42.5}$  erg s $^{-1}$  at  $2.2\mu\text{m}$ . As we show in §3.3, this trend results from the relative host contribution at different bands, which peaks at  $1.2\mu\text{m}$ , decreases sharply in the FUV, and nearly disappears in the X-ray.

The log  $vL_V(\text{FUV})$  vs. log  $L_{\text{bH}\alpha}$  slope is close to unity in the four highest  $L_{\text{bH}\alpha}$  bins, i.e. a linear relation between  $vL_V(\text{FUV})$  and  $L_{\text{bH}\alpha}$ . Since the FUV occurs close to the position of the peak SED in AGN (e.g. Zheng et al. 1997), this linear relation suggests that  $L_{\text{bH}\alpha}$  provides a good estimator of  $L_{\text{bol}}$ . The simple linear relation of  $vL_V(\text{FUV})$  and  $L_{\text{bH}\alpha}$  suggests two independent linear relations. A linear relation of  $vL_V(\text{FUV})$  and the ionizing luminosity, as expected if the mean SED of AGN is luminosity independent. In addition, it requires a linear relation of  $L_{\text{bH}\alpha}$  and the ionizing luminosity, as expected if the mean covering factor of the BLR is also independent of luminosity.

In each panel of Figure 10 we also plot a linear relation, normalized to match the highest  $L_{\text{bH}\alpha}$ -bin. If the AGN SED is indeed luminosity independent, then the deviation from linear relation results from the host contribution, and thus one can use the deviation from the linear relation to estimate the host contribution as a function of wavelength. Note that at  $\lambda \geq 5100\text{\AA}$  the ridge line slope does not approach unity even at the highest  $L_{\text{bH}\alpha}$ -bin, which may suggest the sample does not quite reach the ‘pure’ AGN emission regime (see §3.3).

Greene & Ho (2005) selected a subset of low  $z$  SDSS AGN where the host contribution to the optical emission is minimal, based on the weakness of the Ca II K stellar absorption feature. Their measured  $vL_V(5100\text{\AA})$  vs.  $L_{\text{bH}\alpha}$  relation is plotted in the  $5100\text{\AA}$  panel of Fig. 10, and indeed passes close to the assumed linear relation, normalized at the highest  $L_{\text{bH}\alpha}$ -bin. This strongly supports the host contribution interpretation for the slope changes with decreasing luminosity seen in Fig. 10.

### 3.2.1 The inverse Baldwin relation for $\text{H}\alpha$

We note in passing that the EW of  $\text{H}\alpha$ , given to a good approximation by  $L_{\text{bH}\alpha}/vL_V(7000\text{\AA})$ , decreases with decreasing luminosity due to the increasing relative contribution of the host (Fig. 10,  $7000\text{\AA}$  panel). This corresponds to the ‘inverse Baldwin effect’ of the Balmer lines, noted by Croom et al. (2002) for  $\text{H}\beta$  and  $\text{H}\gamma$  in low luminosity AGN found in the 2dF quasar survey. Unlike the suggestion that the inverse Baldwin effect of the Balmer lines

is an intrinsic property of AGN (Croom et al. 2002), our analysis suggests the intrinsic  $\text{H}\alpha$  EW is luminosity independent, i.e.  $\text{H}\alpha$  shows no Baldwin effect, as was found for  $\text{H}\beta$  in more luminous AGN (Dietrich et al. 2002; cf. Netzer et al. 2004).

### 3.2.2 The $vL_V$ vs. $L_{\text{bH}\alpha}$ for the BQS sample

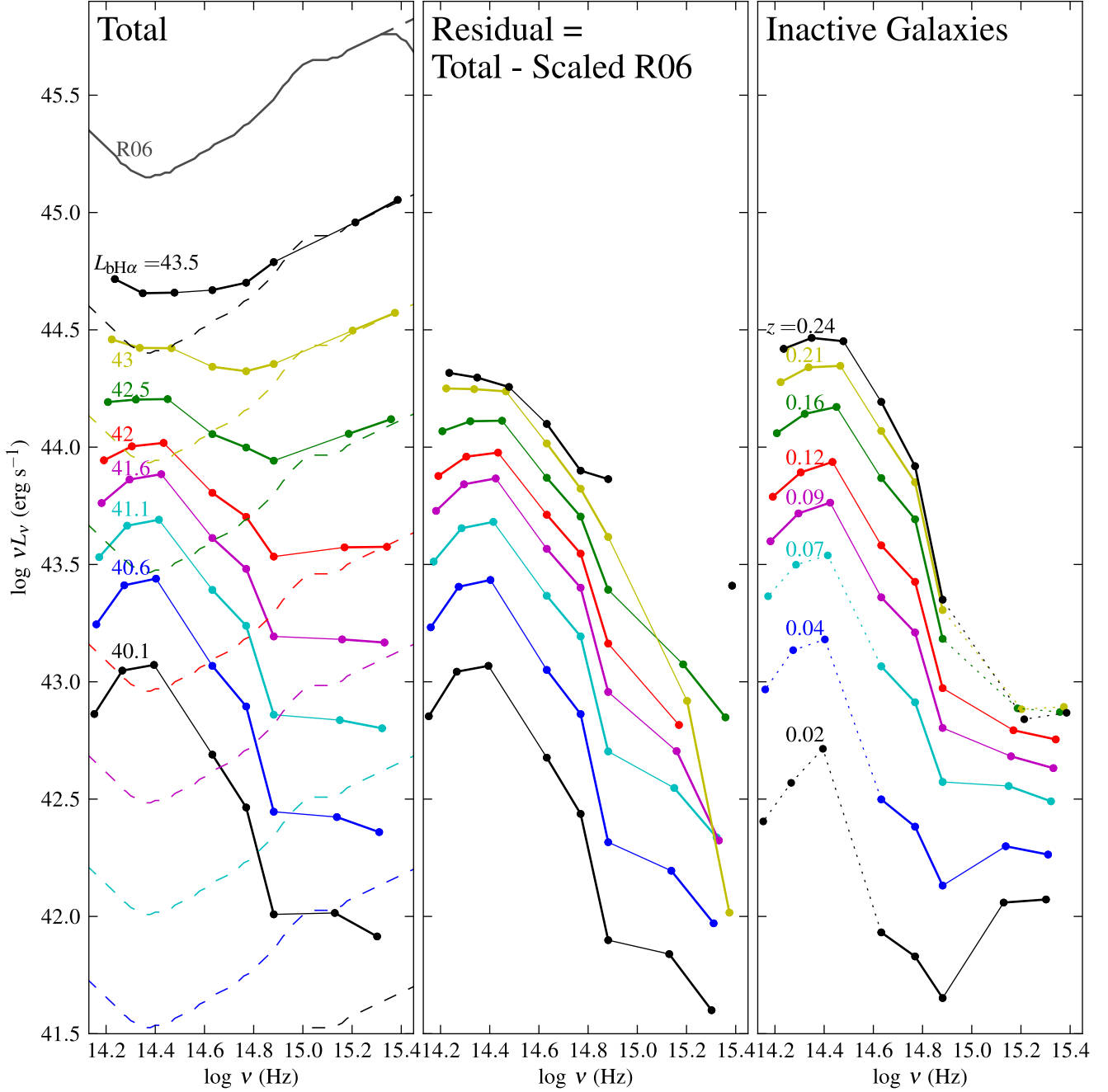
There are 26 objects in the T1 sample that also appear in the BQS sample (Neugebauer et al. 1987). These are marked in Fig. 10. The BQS selection criteria are of point sources with a relatively blue optical colour, in contrast with the T1 selection criteria which are independent of morphology and colour. It is therefore interesting to see how the BQS selection criteria affect their  $L_{\text{bH}\alpha}$  vs.  $vL_V$  correlations. In the  $\lambda \geq 5100\text{\AA}$  bands the BQS objects have a lower  $vL_V$  at a given  $L_{\text{bH}\alpha}$ , and lie closer to the linear relation, in particular for the lowest  $L_{\text{bH}\alpha}$  objects. This offset from the T1 sample is consistent with their selection of a weaker host contribution. At the  $\lambda < 5100\text{\AA}$  bands the BQS objects are offset to higher  $vL_V$  values. The offset increases into the UV, consistent with the optical blue colour selection. In the X-ray band the BQS appear to be a random subset of the T1 sample. Note that the BQS sample objects do not fall outside the observed T1 objects distributions, but are offset towards weaker hosts and stronger UV emission, at a given  $L_{\text{bH}\alpha}$ .

### 3.3 Average SED versus $L_{\text{bH}\alpha}$

Above we have found indirect indications that 1. the mean SED of AGN and the mean covering factor of the BLR are independent of luminosity, 2. the host galaxy contribution flattens the  $vL_V$  vs.  $L_{\text{bH}\alpha}$  relations with decreasing  $L_{\text{bH}\alpha}$ . We explore these points qualitatively in this section. A quantitative analysis is performed in §3.4 – §3.5. Below we form the observed mean AGN SED for the eight  $L_{\text{bH}\alpha}$  bins (Table 2), subtract a mean net AGN SED scaled linearly by  $L_{\text{bH}\alpha}$ , and compare the residual luminosity to the mean SED of a matched sample of inactive galaxies.

Figure 11, left panel, shows the mean SED in the  $2.2\mu\text{m}$  to  $1528\text{\AA}$  region for the eight  $L_{\text{bH}\alpha}$  bins. The integration apertures of 2MASS, SDSS, and GALEX are different, thus the fraction of the host light included is different in different bands. Furthermore, these differences change with  $z$  and depend on the host type. Although the plotted SEDs likely differs from the intrinsic SEDs due to these aperture effects, the difference is irrelevant to our purpose of understanding the SEDs, as long as all objects used in the analysis presented below are similarly affected.

We cannot use our highest  $L_{\text{bH}\alpha}$  SED bin as a template net AGN SED, as the  $vL_V$  vs.  $L_{\text{bH}\alpha}$  relation remains flatter than unity at  $\lambda > 5100\text{\AA}$  even in that bin, indicating possible significant host contribution in those bands. We need a template of more luminous



**Figure 11.** The luminosity dependence of the mean SEDs at  $2.2\mu\text{m}$ – $1528\text{\AA}$ . **(Left panel)** The mean T1 objects observed SEDs for the eight  $L_{\text{bH}\alpha}$  bins (solid lines, see §3.3). Bands from different surveys (different apertures) are connected by a thinner line. Bands with  $< 70\%$  detection are connected by dotted lines. Note the transition from an AGN-dominated UV peaked SED to a galaxy-dominated NIR peaked SED with decreasing  $L_{\text{bH}\alpha}$ . The R06 SED of luminous quasars (top solid gray line), is assumed as the net AGN SED. The dashed lines show the expected net AGN contribution, derived by matching the R06 SED to the UV in the highest  $L_{\text{bH}\alpha}$  bin, and rescaling by the relative  $L_{\text{bH}\alpha}$  of the lower seven bins. The scaled R06 SEDs match the mean SEDs in the UV for the 2<sup>nd</sup>–4<sup>th</sup>  $L_{\text{bH}\alpha}$  bins, as expected from their linear relation (Fig. 10). **(Middle panel)** Residual SEDs derived by subtracting the scaled R06 SED from the total SED (colours match the left panel). For clarity, errors are not shown, though note they are high in the luminous bins, as the residual is a small fraction of the total light. The residuals all appear to have a typical galaxy SED. **(Right panel)** Average SEDs of SDSS inactive galaxies, with  $z$ -distributions matched to the  $z$ -distributions of the T1 objects in each  $L_{\text{bH}\alpha}$  bin (mean  $z$  noted near each curve). The increase of  $L_{\text{host}}$  with  $L_{\text{bH}\alpha}$ , seen in the middle panel, is just a selection effect, as inactive galaxies selected at matched  $z$  show the same trend, inherent to a flux limited sample. Note the significant excess luminosity in the K<sub>s</sub> and H-bands in the most luminous residuals (middle panel), implying that a linear scaling of the net AGN SED is applicable only to the optical-UV region.

AGN, where the host contribution is likely negligible. We therefore use Richards et al. (2006, hereafter R06), which provide a mean AGN SED constructed using SDSS quasars (with Spitzer detections). The R06 template SED is derived from the SDSS photometry, supplemented with photometry from various other surveys, including 2MASS and GALEX. The R06 SED is a factor of  $\sim 5$  more luminous than our highest  $L_{\text{bH}\alpha}$ -bin SED (see Fig. 11), which ensures that host contribution should be lower. In addition, R06 removed the host contribution, which they find is generally  $< 0.1$  dec, even in the J-band. Since the FUV of R06 is mostly derived from SDSS photometry of high  $z$  quasars, it shows a turnover at  $\log \nu \gtrsim 15.32$ , induced by the  $\text{Ly}\alpha$ -forest, which is not present in the T1 sample SEDs. We therefore extrapolate the R06 SED up to  $\log \nu = 15.4$  (the highest relevant rest frame  $\nu$ ) with  $L_\nu \propto \nu^{-0.5}$ , as seen at lower  $\nu$  (see also figure 19 of Trammell et al. 2007).

The R06 SED was scaled to match the NUV data points of the highest  $L_{\text{bH}\alpha}$  bin ( $L_{\text{bH}\alpha} = 10^{43.5} \text{ erg s}^{-1}$ , left panel of Fig. 11). For each of the seven lower luminosity bins, we further rescaled the R06 SED by the mean  $L_{\text{bH}\alpha}$  in each bin, relative to  $L_{\text{bH}\alpha}$  in the highest bin. The scale applied to each bin is listed in Table 2. Note that the  $L_{\text{bH}\alpha}$ -rescaled R06 SEDs match well the FUV data points of the mean SEDs also in the second to fourth  $L_{\text{bH}\alpha}$  bins, as expected from their linear relation with  $L_{\text{bH}\alpha}$  (FUV panel in Fig. 10).

We now subtracted the scaled R06 SED, i.e. the matched ‘pure’ AGN SED, from the mean SED in each bin, and the residuals are plotted in the middle panel of Fig. 11. This residual includes both the host emission, and any change in the mean AGN SED with decreasing  $L_{\text{bH}\alpha}$ . The errors on the residual are high in the luminous bins and in the UV, since the residual is a small fraction of the total light. For clarity sake, error bars are not shown. The possible systematic errors in this method are discussed in §3.4. The residuals have similar overall SEDs, with some small differences at the NIR. They all appear to be similar to a typical galaxy SED, which peaks at  $\log \nu = 14.4$ . The residuals imply that the mean host galaxy luminosity increases with the mean  $L_{\text{bH}\alpha}$  luminosity, which suggests a tight relation between the AGN luminosity and the host luminosity. However, as we show below, this apparently tight relation is just a selection effect.

To compare the residuals to the mean SED of typical galaxies, we formed a matched- $z$  sample of 298 SDSS galaxies classified by B04 as inactive (excluding ‘type 2 AGN’ or ‘composites’, see §3.1.4). The galaxies were drawn from the GALEX Deep Imaging Survey region, to maximize UV detections, and thus minimize bias in terms of their UV emission. Their SEDs were formed in a manner identical to the T1 sample. For each  $L_{\text{bH}\alpha}$ -bin of the T1 sample we formed a  $z$ -matched sample of inactive galaxies in the following manner. First, the inactive galaxies are divided into eight  $z$  bins, evenly spaced in  $z^{1/2}$ , and the mean-SED of each  $z$ -bin is computed, forming ‘ $z$ -specific’ SEDs. Then, the  $z$ -distribution of each AGN  $L_{\text{bH}\alpha}$ -bin is measured. Finally, the ‘ $z$ -specific’ SEDs are convolved with the different  $z$ -distribution functions. Note that only a fraction of the high luminosity (and high  $z$ ) AGN are resolved (Table 2), while all inactive galaxies are resolved by definition (§2.1). Therefore, the high  $z$  AGN are subject to somewhat different selection effects from inactive galaxies with the same  $z$ .

The mean-SEDs of the  $z$ -matched inactive galaxies, are shown in the right panel of Figure 11. The convolved detection fraction is  $< 70\%$  at high  $z$  in the FUV, and at low  $z$  in the NIR, so the SEDs may be somewhat biased in these regions. Except in the lowest  $z$  bins, the residuals largely resemble the inactive galaxies, both in shape and in luminosity. The trend of increasing host luminosity

with the AGN luminosity, seen in the middle panel, just reflects the fact that higher luminosity AGN reside at higher  $z$ . Since most of the T1 objects are selected from the SDSS sample of extended objects (galaxies), and the SDSS sample is flux limited, the higher  $z$  AGN must reside in higher luminosity host galaxies. Now, for the unresolved AGN, which dominate at the highest  $L_{\text{bH}\alpha}$  (see Table 2), the hosts are not forced to be more luminous to be detectable at higher  $z$ , and indeed the inactive galaxies are somewhat more luminous than the AGN hosts at the highest  $L_{\text{bH}\alpha}$  bins. Unresolved AGN flag the position of their host galaxies, galaxies which would have not necessarily passed the flux limit of the SDSS survey.

The subtraction of the scaled R06 AGN SED leads to a residual SED which is generally consistent with the mean SED of inactive galaxies in the middle and high  $L_{\text{bH}\alpha}$  bins, indicating the linear scaling of the R06 SED with  $L_{\text{bH}\alpha}$  is roughly correct. At the lowest  $L_{\text{bH}\alpha}$  bin the inactive galaxies show excess UV emission. This is unlikely to be an AGN SED effect, as the host dominates the emission at all bands (Fig. 11, left panel), and likely reflects the difference in the mean AGN host galaxy (NEG) and the mean SDSS inactive galaxy at low  $z$  (SFG, see Fig. 9). There is also significant excess luminosity in the  $K_s$  and H-bands in the most luminous residuals. This could be due to a lower dust covering factor with increasing luminosity, as observed by Maiolino et al. (2007) and Treister et al. (2008), and thus weaker dust emission in the R06 template compared to the T1 AGN. We do not attempt to quantify this relation or pursue it further, yet we note that the simple scaling law of a fixed AGN SED does not apply in the near IR. Note that the X-ray emission also does not scale linearly with  $L_{\text{bH}\alpha}$  either (Fig. 10 and §3.6). Therefore, we emphasize the applicability of this simple scaling law is limited only to AGN optical-UV emission, the region most likely dominated by the accretion disc emission.

### 3.4 The AGN SED scaling with $L_{\text{bH}\alpha}$

We express the net AGN  $\nu L_\nu$  ( $\equiv L_{\text{AGN}}$ ) by scaling the R06 SED by  $L_{\text{bH}\alpha}$ :

$$L_{\text{AGN}}(\lambda) \equiv (L_{\text{R06}}(\lambda) \pm 10\%) \times \left( \frac{L_{\text{bH}\alpha}}{10^{44.2} \text{ erg s}^{-1}} \right)^\beta \quad (4)$$

The normalization of  $L_{\text{bH}\alpha}$ ,  $10^{44.2} \text{ erg s}^{-1}$ , corresponds to  $L_{\text{bH}\alpha} = 10^{43.5} \text{ erg s}^{-1}$  of the highest luminosity bin, times the difference of  $10^{0.7}$  in  $\nu L_\nu(\text{NUV})$  between this bin and the R06 SED, as described in the previous section. The scaling index is denoted by  $\beta$ , where a  $\beta$  independent of  $\lambda$  indicates that the mean SED shape remains fixed with  $L_{\text{bH}\alpha}$ . We use a 10% systematic error on the scaled R06 luminosity to get an estimate of the implied errors in the residual fluxes by systematic errors of a given level.

#### 3.4.1 Constraining $\beta$

In §3.3, we found indications that  $\beta \approx 1$ . The left panels in Figure 12 provide additional support for this linear scaling law. It compares the implied host luminosity ( $L_{\text{host}} \equiv \text{mean } \nu L_\nu - \text{mean } L_{\text{AGN}}$ ) at different  $L_{\text{bH}\alpha}$  bins, for different  $\beta$ , to the mean luminosities of SDSS type 2 AGN with the same  $z$  distribution. The luminosity is presented in two optical bands: the SDSS  $z$  band (8932Å) and the SDSS  $u$  band (3551Å). The  $z$  band is selected as it follows the stellar mass. The SDSS  $u$  band is a measure of the hotter, more massive stars, and thus a measure of the more recent star formation rate (SFR). Going further to the UV bands probes hotter and more massive stars, and provides a measure of the yet more recent SFR.



However, given the little residual emission left in the UV, the uncertainty can be large. For all objects, we use the  $u$  band and the  $z$  band photometric luminosities in the 3'' SDSS fibre aperture, fixed by 0.35 mag, the mean offset for a PSF (Adelman-McCarthy et al. 2008, see also §2.7). Magnitudes are converted to physical units using the SDSS-AB offsets in Abazajian et al. (2004). We use only the 2 766 objects from the T1 sample drawn from the SDSS sample of galaxies, which have the same selection criteria for spectroscopy as the type 2 AGN (§2.1). We use only type 2 AGN (classified by B04, see §3.1.4) that pass the S/N and bad pixel criteria applied to the T1 sample (§2.1).

Fig. 12 presents  $L_{\text{host}}$  derived for  $\beta = 0.9, 1.0$  and  $1.2$ . The three left panels present the mean  $L_{\text{host}}(u \text{ band})$ ,  $L_{\text{host}}(z \text{ band})$  and  $u - z(\text{host})$  colour for the T1 and for  $z$  matched type 2 AGN (the mean  $z$  for each  $L_{\text{bH}\alpha}$  bin is marked at the top of the figure). We also plot the relations for inactive galaxies here selected from the complete SDSS survey (not limited to the deep survey field of GALEX). The error bars on the  $\beta = 1$  line depict the effect of a 10% systematic error in the scaling of the R06 SED (eq. 4), added to the error in the mean. The possible systematic error has a significant effect at the highest  $L_{\text{bH}\alpha}$  bins, where  $L_{\text{host}}$  becomes a small fraction of the total luminosity, in particular in the  $u$  band. At the highest  $\log L_{\text{bH}\alpha} = 43$  bin, the scaled R06 becomes higher than the mean  $u$  band luminosity, leading to an implied negative  $L_{\text{host}}$ , and only the upper limit from the systematic error is shown. At lower  $L_{\text{bH}\alpha}$  bins the scaled R06 SED becomes much weaker than the observed mean SED (left column of Fig. 11), and the host SED nearly equals the observed mean SED.

As Fig. 12 clearly shows, a  $\beta = 1$  leads to a good match of the T1 host galaxies to the host galaxies of type 2 AGN, as expected based on AGN unification. Values of  $\beta < 0.9$  and  $\beta > 1.2$  in the  $u$  band lead to highly discrepant hosts. The  $z$  band is more host dominated, and less sensitive to the scaled AGN subtraction, so it does not provide a strong constraint on  $\beta$ . It is remarkable that the simplest scaling,  $\beta = 1$ , i.e. that the net AGN SED maintains a fixed shape, which scales linearly with  $L_{\text{bH}\alpha}$ , leads to the closest match of the implied type 1 host to the observed type 2 host. The small differences in overall luminosity between type 1 hosts and type 2 AGN at low  $L_{\text{bH}\alpha}$  (low  $z$ ) cannot be due to a different  $\beta$ , as the AGN light is too weak. They could either be a result of the different selections used to create the type 1 and type 2 samples, or some modification of the unified model at low luminosities.

### 3.4.2 The mean net AGN SED

Combining the  $\beta = 1$  deduced from the  $u$  and  $z$  bands (Fig. 12) with the linear relation of  $vL_V(\text{FUV})$  and  $L_{\text{bH}\alpha}$  (Fig. 10), we get that three bands in the 1500Å – 9000Å range scale linearly with  $L_{\text{bH}\alpha}$ , implying that the optical-UV net AGN SED shape is independent of luminosity. This result is consistent with the finding of Bentz et al. (2009) for a sample of 35 reverberation mapped AGN, where they found that the dependence of the BLR radius on AGN optical luminosities, corrected for the host emission, is the same as the dependence on UV luminosities (Kaspi et al. 2005), which implies that the net AGN optical luminosity scales linearly with their UV luminosity.

Below we list the ratios of AGN continuum emission in different rest-frame wavelengths to  $L_{\text{bH}\alpha}$ , as implied by the linear re-

lation and by the R06 SED:

$$L_{\text{AGN}} \begin{pmatrix} 1450\text{\AA} \\ 2500\text{\AA} \\ 3940\text{\AA} \\ 5100\text{\AA} \\ 6166\text{\AA} \\ 7000\text{\AA} \\ 1.2\mu\text{m} \end{pmatrix} = \begin{pmatrix} 33^{+2.4}_{-2.4} \\ 26 \\ 17 \\ 14 \\ 12 \\ 11 \\ 8.5 \end{pmatrix} \times L_{\text{bH}\alpha} \quad (5)$$

The implied mean EW of the broad H $\alpha$  is 570Å. The dispersion is derived from the scatter around the linear relation between  $\log vL_V(\text{FUV})$  and  $\log L_{\text{bH}\alpha}$ , in the four luminous bins (Fig. 10). We do not attempt to measure the dispersion in the  $L_{\text{bH}\alpha}$  vs.  $L_{\text{AGN}}(\lambda)$  relations at optical  $\lambda$ , due to the unknown dispersion in the host contribution. However, the dispersion at optical  $\lambda$  is smaller than in the UV (Fig. 10), despite the additional dispersion due to the host contribution. Therefore, for  $\lambda > \text{FUV}$ , we treat the noted scatter of 2.4 as an upper limit on the scatter of  $L_{\text{AGN}}(\lambda)$  vs.  $L_{\text{bH}\alpha}$ .

Using the conversion of R06 from  $vL_V$  at various bands to the bolometric luminosity (see their figure 12),  $L_{\text{bol}}$ , we get

$$L_{\text{bol}} = 130^{+2.4}_{-2.4} \times L_{\text{bH}\alpha} \quad (6)$$

## 3.5 The AGN host

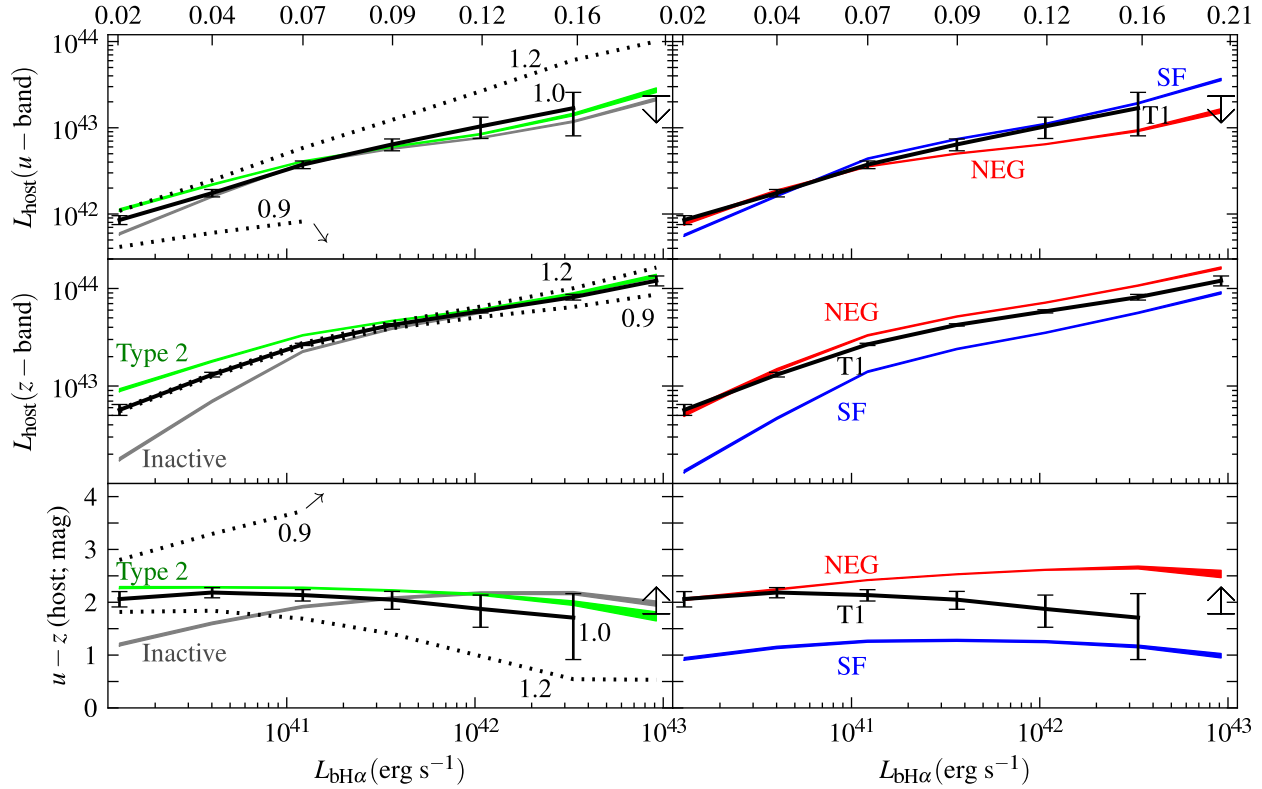
### 3.5.1 Host mass

As discussed above (§3.4, and Fig. 11), the apparent correlation of  $L_{\text{host}}$  and  $L_{\text{bH}\alpha}$ , is most likely just a selection effect, as the mean luminosity of inactive galaxies, of similar  $z$  distribution, rises similarly with  $z$ . This similarity strongly suggests that the apparent correlation between  $L_{\text{bH}\alpha}$  and the AGN  $L_{\text{host}}$  is induced by the correlation of both quantities with  $z$ . To test if the  $L_{\text{host}}$  versus  $L_{\text{bH}\alpha}$  correlation is significant, we use the T1 sample presented in Fig. 12 (selected from the galaxy catalogue), to perform a partial correlation analysis. The T1  $L_{\text{host}}(z \text{ band})$  is indeed strongly correlated with  $L_{\text{bH}\alpha}$ , with a Pearson correlation coefficient  $P_r(L_{\text{host}}, L_{\text{bH}\alpha}) = 0.66$ . However, both quantities are indeed strongly correlated with  $z$ , with  $P_r(L_{\text{host}}, z) = 0.82$ , and  $P_r(L_{\text{bH}\alpha}, z) = 0.74^8$ , and the partial correlation of  $L_{\text{host}}(z \text{ band})$  versus  $L_{\text{bH}\alpha}$ , at a fixed  $z$ , is only  $P_r(L_{\text{host}}, L_{\text{bH}\alpha})|_z = 0.14$ . Furthermore, a fit of the residuals in the  $L_{\text{host}}(z \text{ band}) - z$  relation versus the residuals in the  $L_{\text{bH}\alpha} - z$  relation gives  $L_{\text{host}}(z \text{ band}) \propto L_{\text{bH}\alpha}^{0.07}$ , which again demonstrates there is no significant dependence of  $L_{\text{host}}(z \text{ band})$  on  $L_{\text{bH}\alpha}$ .

Note that  $L_{\text{host}}$  and  $L_{\text{AGN}}$  are not entirely unrelated, as one expects a lower limit to  $L_{\text{host}}$  at a given  $L_{\text{AGN}}$ , in order not to exceed  $L_{\text{Edd}}$ . But the T1 subsample of extended objects excludes AGN close to  $L_{\text{Edd}}$  (§3.1.2, Fig. 8).

As discussed above (§3.1.4), the T1  $L_{\text{host}}$  distribution follows the NEG luminosity distribution (Fig. 9). We therefore break the group of inactive galaxies (Fig. 12, left panels) to its two subgroups of NEG and SFG (Fig. 12, right panels). At low  $L_{\text{bH}\alpha}$ ,  $L_{\text{host}}(z \text{ band})$  follow well NEG. This similarity is consistent with the tendency of low luminosity AGN to reside in early type galaxies (Ho et al. 1997b). With increasing  $L_{\text{host}}$ , the T1 hosts become less luminous in the  $z$  band than NEG, and at the highest luminosities the T1 hosts are intermediate between NEG and SFG.

<sup>8</sup> The values of  $P_r$  are calculated using  $\log L_{\text{bH}\alpha}$ ,  $\log L_{\text{host}}$  and  $\log z$ .



**Figure 12. (Left panels)** A comparison of the implied host luminosity (total - scaled R06), for different scaling laws, with  $z$ -matched type 2 AGN, in the  $u$  and  $z$  bands. The residuals vs.  $L_{bH\alpha}$  are shown as solid/dotted lines for linear/non-linear scaling laws. The scaling index,  $\beta$  (eq. 4), is noted near each line. Error bars are the sum of a 10% systematic error (eq. 4) and the error in the mean. The mean  $z$  is noted on top. Also shown are  $z$ -matched inactive galaxies. In the  $u$  band,  $\beta < 0.9$  is ruled out, as the implied residual is negative for  $L_{bH\alpha} > 10^{41}$  erg s $^{-1}$ . It is remarkable that the simplest scaling,  $\beta = 1$  in both bands, leads to the closest match of the mean type 1 host to the mean type 2 host. Together with the linear  $vL_{\nu}(\text{UV})$ - $L_{bH\alpha}$  relationship at high luminosities (Fig. 10), this implies the mean optical-UV net AGN SED shape is fixed and scales linearly with  $L_{bH\alpha}$ . The small offset of T1 and type 2 at low  $L_{bH\alpha}$  cannot be explained by  $\beta \neq 1$  as the AGN emission is a small fraction of the total light. **(Right panels)** A comparison of the residual implied from the  $\beta = 1$  scaling law, with inactive galaxies (left panel) now broken to NEG and SFG. At low  $L_{bH\alpha}$ , hosts of T1 AGN resemble NEG, both in mass ( $z$  band) and in specific SFR (i.e. colour), consistent with the tendency of low luminosity AGN to reside in elliptical hosts. With increasing  $L_{bH\alpha}$ , the T1 hosts become less luminous in the  $z$  band and bluer than the average NEG. The implied SFR is noted in Table 3 and §3.5.

### 3.5.2 Host color

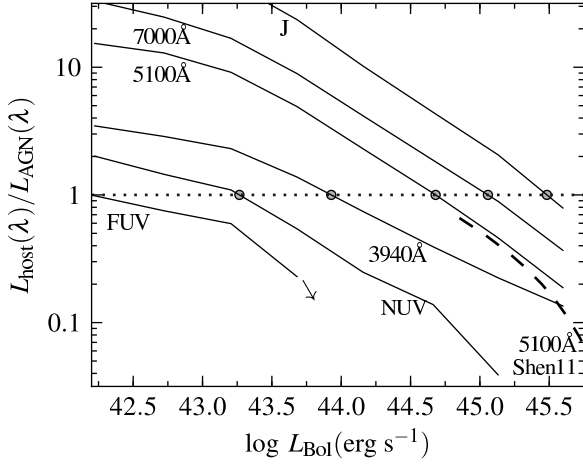
The  $u - z$  colour of the T1 host follows well NEG at the lowest  $L_{bH\alpha}$ . With increasing  $L_{bH\alpha}$  the T1 hosts become bluer, and are intermediate between SFG and NEG at the highest  $L_{bH\alpha}$ . Kauffmann et al. (2003a) find that low luminosity type 2 AGN have NEG colours, while high luminosity type 2 AGN tend to have colours intermediate between NEG and SF galaxies. The results above indicate that type 1 AGN have the same hosts as type 2 AGN. Note that our selection of type 1 AGN is independent of the distinction between SFG and NEG, which is based on narrow lines. In contrast, a type 2 AGN sample, selected based on narrow line ratios, excludes a SF-dominated host galaxy. Although, as shown by Kauffmann et al. (2003a, fig. 6 there), only the lowest luminosity AGN can be dominated by the host SF line emission.

In Table 3, we derive the SFR and specific SFR for each  $L_{bH\alpha}$  bin, using the rest-frame  $u$  band luminosity and  $u - r$  colour of the mean hosts. We calculate the rest frame  $u$  band and  $r$  band luminosities using a power law interpolation between adjacent observed bands. The third column lists  $b_{300}(u-r)$  (Blanton & Roweis 2007), the ratio of stars formed in the last 300 million years to the total star formation history of the galaxy, based on SED modeling of SDSS galaxies using the stellar models of Bruzual & Charlot (2003). The

$\log L_{bH\alpha}$	$z$	$\log b_{300}$	SFR ( $M_{\odot} \text{ yr}^{-1}$ )
40.1	0.02	$-2.89^{+0.16}_{-0.17}$	$0.11^{+0.01}_{-0.01}$
40.6	0.04	$-2.92^{+0.13}_{-0.14}$	$0.24^{+0.02}_{-0.02}$
41.1	0.07	$-2.73^{+0.12}_{-0.13}$	$0.60^{+0.05}_{-0.05}$
41.6	0.09	$-2.56^{+0.16}_{-0.18}$	$1.1^{+0.1}_{-0.1}$
42.0	0.12	$-2.38^{+0.25}_{-0.32}$	$1.9^{+0.4}_{-0.4}$
42.5	0.15	$-2.28^{+0.40}_{-0.65}$	$3.5^{+1.1}_{-1.3}$

**Table 3.** The dependence of mean aperture SFR (within the SDSS fibre) on AGN luminosity. Col. 3 lists the mean ratio of stars formed in the last  $3 \times 10^8$  years compared to the total star formation of the galaxy, derived from rest-frame  $u - r$  colour (Blanton & Roweis 2007). Col. 4 lists the SFR derived from  $u$  band luminosity (Moustakas et al. 2006). Both measures are based on the mean T1 AGN, after subtracting the scaled R06 SED (Fig. 11). Errors are the sum of a 10% systematic error (eq. 4) and the error in the mean.

fourth column lists the current SFR, based on the correlation of the  $u$  band fibre luminosity with  $L_{H\alpha}$  in SDSS SFG (Moustakas et al. 2006). These estimates are subject to a host of systematic effects, such as reddening, the time dependence of the star formation, and obviously the estimated host luminosity.



**Figure 13.** The mean host to AGN luminosity ratio as a function of AGN bolometric luminosity at different rest-frame bands, based on the decomposition in Fig. 11. The small gray circles mark the transition luminosity from host dominance to AGN dominance. In the J band the host dominates up to  $L_{\text{bol}} = 10^{45.5} \text{ erg s}^{-1}$ , well into the quasar regime, while in the FUV the AGN dominates already at  $L_{\text{bol}} > 10^{42.5} \text{ erg s}^{-1}$ . The ratio at  $\lambda = 5100\text{\AA}$  is consistent with the Shen et al. (2011) result (dashed line) which is based on the QCV sample. Note that the results are relevant for the specific apertures used by GALEX, SDSS and 2MASS.

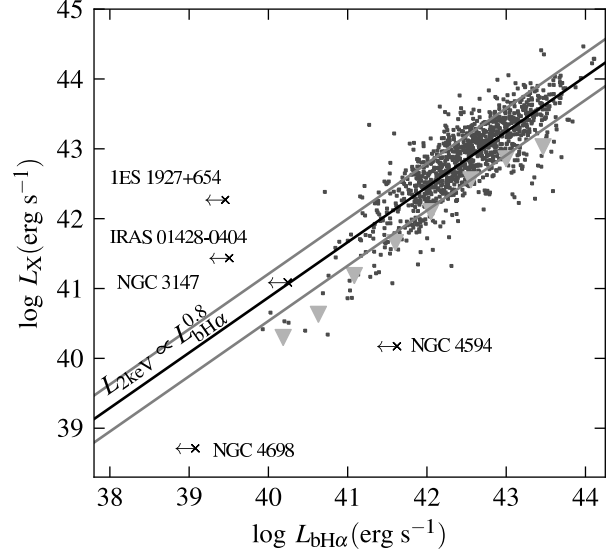
### 3.5.3 The mean host/AGN luminosity ratio

The results described above provide a practical measure of the average host/AGN luminosity ratio in various bands, as a function of the AGN luminosity for SDSS, GALEX, and 2MASS. Figure 13 plots the host luminosity (i.e. residual luminosity) to the net AGN luminosity (i.e. the scaled R06 SED emission), as a function of  $L_{\text{bol}}$  (eq. 6) for different rest-frame bands. In the J band the host dominates, on average, up to  $\log L_{\text{bol}} = 45.5$ , well into the quasar-scale emission, including the highest  $L_{\text{bH}\alpha}$ -bin in our analysis. At  $5100\text{\AA}$  the host dominates at  $\log L_{\text{bol}} \leq 44.7$ , consistent with the relation obtained by Shen et al. (2011) for the SDSS quasar sample. As expected, in the FUV the AGN becomes more dominant already at  $\log L_{\text{bol}} > 42.5$ . The plotted ratios are strongly dependent on the apertures used in each band. Also, the dispersion of  $L_{\text{host}}/L_{\text{AGN}}$  in each  $L_{\text{bH}\alpha}$  bin is significant, so individual objects may differ significantly. The plotted ratio only serves as a practical guide for the AGN catalogues of SDSS, 2MASS and GALEX samples used here.

### 3.6 Unobscured type 2 AGN based on the $L_X$ versus $L_{\text{bH}\alpha}$ relation

Unobscured type 2 AGN, or the so-called ‘true type 2’ AGN, are AGN in which the BLR is not observed because it does not exist, as opposed to standard type 2 AGN in which the BLR is obscured. Candidates are often found by looking for low luminosity AGN, detected through a nuclear X-ray point source, that do not show a sign of a BLR in their optical continuum, but their X-ray spectrum does not indicate obscuration (Tran et al. 2011, Shi et al. 2010, Rigby et al. 2006, Trump et al. 2009). Below we describe how the  $L_X(\equiv \nu L_\nu(2 \text{ keV}))$  versus  $L_{\text{bH}\alpha}$  relation can be used to test whether the lack of detection of broad lines is indeed significant, and justifies the true type 2 identification.

Our X-ray band detection rate for the complete T1 sample is



**Figure 14.** The  $L_X$  vs.  $L_{\text{bH}\alpha}$  relation as a probe for true type 2 AGN. The plot includes only T1 objects with  $F_{\text{bH}\alpha} > 10^{-13.5} \text{ erg s}^{-1} \text{ cm}^{-2}$ , where  $> 50\%$  have ROSAT detections (Fig. 3). The non detections are marked by triangles, which show the mean upper limits, averaged in half decade wide  $L_{\text{bH}\alpha}$  bins. The best power law fit for the X-ray detections is shown as a black line, with gray lines indicating the dispersion. The left pointing arrows mark five true type 2 candidates from Tran et al. (2011) and Shi et al. (2010). These objects have an apparently unobscured X-ray emission, but no broad  $\text{H}\alpha$  detection. Only IES 1927+654 and IRAS 01428-0404 have upper limits on  $L_{\text{bH}\alpha}$  which are significantly lower than expected based on their  $L_X$ , and are therefore potentially ‘true type 2’ AGN. In NGC 4594 and NGC 4698 the expected  $L_{\text{bH}\alpha}$ , based on their  $L_X$ , is a factor of  $\sim 100$  below the upper limit on  $L_{\text{bH}\alpha}$ , and in NGC 3147 50% of the objects are expected to be below the  $L_{\text{bH}\alpha}$  upper limit. The upper limit on  $L_{\text{bH}\alpha}$  depends on the expected  $\Delta\nu$ , which can be derived from the  $M_{\text{BH}}$  and  $L/L_{\text{Edd}}$  estimates for each object.

43%. The detection rate depends on  $L_{\text{bH}\alpha}$ , and becomes as low as 20% in the lowest  $L_{\text{bH}\alpha}$  bins (Table 2). Therefore, in the following analysis of the  $L_X$  versus  $L_{\text{bH}\alpha}$  relation we utilize only objects with  $F_{\text{bH}\alpha} > 10^{-13.5} \text{ erg s}^{-1} \text{ cm}^{-2}$ , which have an X-ray detection rate of 61% (see Fig. 3). Figure 14 presents the  $L_X$  versus  $L_{\text{bH}\alpha}$  relation. The  $L_X$  of inactive galaxies is generally  $< 10^{42} \text{ erg s}^{-1}$  (Bauer et al. 2004), so host contribution should not significantly affect our analysis.

A least squares minimization yields

$$\log L_{X,42} = (0.79 \pm 0.02) \log L_{\text{bH}\alpha,42} + (0.45 \pm 0.01) \quad (7)$$

$$\alpha_{\text{ox}} = (-0.09 \pm 0.01) \times \log l_{2500\text{\AA},29} - 1.42 \quad (8)$$

where  $\alpha_{\text{ox}} \equiv 0.38 \times \log(l_{2 \text{ keV}}/l_{2500\text{\AA}})$ ,  $L_{\text{bH}\alpha}$  and  $L_X$  are in units of  $10^{42} \text{ erg s}^{-1}$  and  $l_{2500\text{\AA}}$  in  $10^{29} \text{ erg s}^{-1} \text{ Hz}^{-1}$ . Equation 8 is similar to eq. 2 in Steffen et al. (2006), where  $\alpha_{\text{ox}} \propto \log l_{2500\text{\AA}}^{(-0.137 \pm 0.008)}$ . The fit relation (eq. 7) is depicted in Fig. 14. For each  $L_{\text{bH}\alpha}$ -bin, a triangle marks the mean upper limit on  $L_X$  of the T1 objects without ROSAT detections.

Eq. 7 provides a useful relation for the search for the putative unobscured type 2 AGN. Below we compare the upper limit on  $L_{\text{bH}\alpha}$  for a number of potentially ‘true type 2’ AGN presented by Tran et al. (2011) and Shi et al. (2010) with their predicted  $L_{\text{bH}\alpha}$  based on their observed  $L_X$ .

For each object we set an upper limit on the broad  $H\alpha$  flux as 10% of the continuum flux density near  $H\alpha$ , multiplied by the expected  $\Delta v$ . The expected  $\Delta v$  is estimated from  $M_{BH}$  and  $L/L_{Edd}$ , by applying eqs. 2 and 3. We use the three objects in Tran et al. (2011) where estimates of  $M_{BH}$  and  $L/L_{Edd}$  are provided, and also two objects from Shi et al. (2010), NGC 4594 and IRAS 01428-0404, where optical spectra and  $M_{BH}$  estimates are provided. These five objects have a measured  $L_X$ , and we use eq. 7 to derive the expected  $L_{bH\alpha}$  based on their observed  $L_X$ , and compare it to the upper limit on  $L_{bH\alpha}$  derived above. The five objects are plotted in Fig. 14 (left-pointing arrows). Three of the potentially true type 2 candidates, NGC 4594, NGC 4698, and NGC 3147 have upper limits on  $L_{bH\alpha}$  that can accommodate the expected values of  $L_{bH\alpha}$  based on the extension of eq. 7 to lower luminosities. Thus, the absence of a broad  $H\alpha$  in these objects is not significant. A high angular resolution spectrum is required to exclude the strong host contribution near  $7000\text{\AA}$ , and determine if the expected weak  $H\alpha$  is indeed missing. In the other two objects, IRAS 01428-0404 and IES 1927+654, the expected  $L_{bH\alpha}$  is well above the upper limits. The deficiency in  $L_{bH\alpha}$  is significant at the levels of  $4\sigma$  and  $7\sigma$ , respectively<sup>9</sup>, and these two objects appear to be true type 2 AGN.

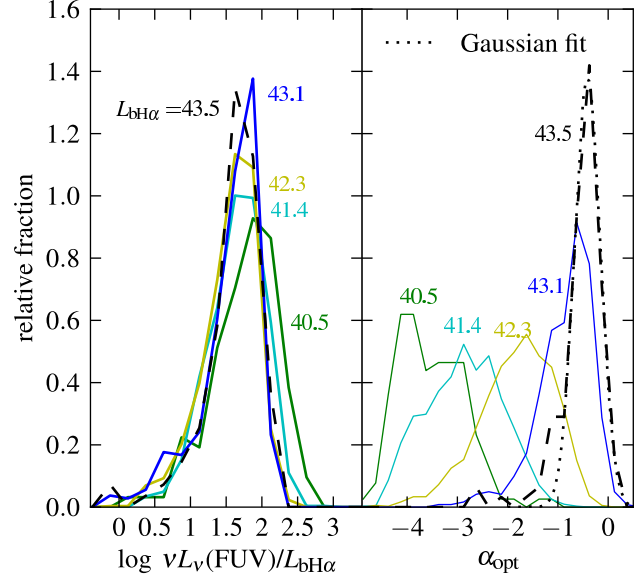
An estimate of the expected  $L_{bH\alpha}$  can also be derived directly from  $L_{bol}$ , using eq. 6. However, the estimated  $L_{bol}$  in Tran et al. and Shi et al. is based on  $L_{[O III]}$ , and its relation with AGN luminosity has significant dispersion (Baskin & Laor 2005, Caccianiga & Severgnini 2011). This large dispersion is less inhibiting in the estimation of  $\Delta v$ , used to derive the upper limit on  $L_{bH\alpha}$ , since  $\Delta v \propto L_{bol}^{-1/4}$  when  $m$  is known (eq. 2). An analysis of the relation between  $L_{[O III]}$  and  $L_{bH\alpha}$  in the T1 sample, and its implication on the identification of true type 2 AGN, will be performed in a following paper.

### 3.7 The SED dispersion

Above we have shown that the mean SED of broad line AGN is well reproduced by the sum of the mean SED of luminous quasars, scaled down by  $L_{bH\alpha}$ , plus the mean SED of inactive galaxies (NEG-like at low  $L_{bH\alpha}$ , and bluer with increasing  $L_{bH\alpha}$ ). However, what is the dispersion in the shapes of individual SEDs? Below we provide some measures of the distribution of the SEDs, and discuss its possible origin.

#### 3.7.1 The distribution of $vL_V(FUV)/L_{bH\alpha}$ and $\alpha_{opt}$

Figure 15, left panel, presents the distribution of  $vL_V(FUV)/L_{bH\alpha}$  in four decade-wide bins of  $L_{bH\alpha}$  (coloured solid lines, each bin in Fig. 15 is composed of two consecutive  $L_{bH\alpha}$  bins from Table 2). The distributions for  $L_{bH\alpha} \geq 10^{41.4} \text{ erg s}^{-1}$  show a narrow peak at  $vL_V(FUV)/L_{bH\alpha} = 45 - 75$ , compared to the mean value of 33 (eq. 5). There is an extended tail towards lower values, which is similar in all three bins. The amplitude of the narrow peak drops with decreasing mean  $L_{bH\alpha}$ . In the lowest bin,  $L_{bH\alpha} = 10^{40.5} \text{ erg s}^{-1}$ , the peak becomes broader and is shifted to higher  $vL_V(FUV)/L_{bH\alpha}$  values. As found above (Fig. 10 FUV panel), this bin is affected by significant host contribution to the FUV band, which produces



**Figure 15. (Left panel)** The distribution of the  $vL_V(FUV)/L_{bH\alpha}$  values as a function of  $L_{bH\alpha}$ . Four  $L_{bH\alpha}$  bins, a decade wide, are plotted (solid lines). The mean  $L_{bH\alpha}$  is noted near each distribution. The upper half of the upper bin, with  $43.3 < \log L_{bH\alpha} < 43.8$ , is also shown (dashed line). The distributions for  $\log L_{bH\alpha} > 40.5$  have a very similar shape. All distributions are asymmetric, with an extended tail towards low  $vL_V(FUV)/L_{bH\alpha}$  values, which may be produced by reddening. In the  $\log L_{bH\alpha} = 40.5$  bin, host contribution of the UV broadens the distribution and increases the mean value. **(Right panel)** The distribution of optical slope, also binned by  $L_{bH\alpha}$ . The host emission makes  $\alpha_{opt}$  redder with decreasing AGN luminosity. The upper half of the highest luminosity bin is least affected by the host contribution, and is well fit by a Gaussian (dotted line) and an extended red tail, which may also be produced by reddening. There is therefore a small dispersion in  $vL_V(FUV)/L_{bH\alpha}$  and  $\alpha_{opt}$  in the net AGN emission, based on the objects least affected by the host contribution.

the offset to higher  $vL_V(FUV)/L_{bH\alpha}$  values, and a larger dispersion. The higher  $L_{bH\alpha}$  bins are not affected by the host contributions, and the observed distribution of  $vL_V(FUV)/L_{bH\alpha}$  values is an intrinsic AGN property. The distribution of these three bins is similar, despite being selected differently by the SDSS, where the  $\log L_{bH\alpha} = 43$  objects are mainly point sources, selected by their non-stellar colours, while lower  $L_{bH\alpha}$  objects are selected to have extended morphologies (§2.1 and Table 2). Since  $H\alpha$  is generally powered by recombination, it provides a measure of the ionizing photons intercepted by the BLR. Thus, the narrow peak suggests there is a small dispersion in  $L_{ionizing}/vL_V(FUV)$ , and a small dispersion in the covering factor of the BLR. The tail towards lower  $vL_V(FUV)/L_{bH\alpha}$  values may be due to reddening, as further discussed below based on additional indicators.

Figure 15, right panel, shows the distribution of the SDSS based optical slope  $\alpha_{opt} \equiv 4 \times \log(l_V(3940\text{\AA})/l_V(7000\text{\AA}))$  as a function of  $L_{bH\alpha}$ . The  $\alpha_{opt}$  distribution moves to redder values with decreasing  $L_{bH\alpha}$ , reflecting the increasing amount of host contribution. The highest  $L_{bH\alpha}$  bin (43.1) is least affected by the host contribution, and to further decrease the effect of the host contribution we show the upper half of the  $\log L_{bH\alpha} = 43.1$  bin, where  $\log L_{bH\alpha} = 43.5$  ( $43.3 < \log L_{bH\alpha} < 43.8$ ). The distribution is well characterized as a Gaussian (see plot) with a red tail excess. This distribution is similar to the slope distribution found in luminous quasars by Richards et al. (2003), measured between  $1450\text{\AA}$

<sup>9</sup> Our calculation of the dispersion assumes  $L_{bH\alpha}$  is the independent variable, since the T1 sample is selected by the broad  $H\alpha$ . When estimating  $L_{bH\alpha}$  from  $L_X$  one should use an X-ray selected sample and treat  $L_X$  as the independent variable.

and 4040Å of composite spectra. Their slope distribution peaks between  $-0.54$  and  $-0.41$  (table 1 there), while 6% of their objects are in a red tail, attributed to a reddening effect. Here, the peak of the distribution of the  $L_{\text{bH}\alpha} \sim 10^{43.5} \text{ erg s}^{-1}$  objects is at  $\alpha_{\text{opt}} = -0.39$ , and 12% of the objects are in the red tail with  $\alpha_{\text{opt}} < -1$ . This similarity with Richards et al. (2003) indicates that the AGN slope distribution does not change significantly with luminosity, and that the host effect on the scatter in  $\alpha_{\text{opt}}$  is negligible at  $\log L_{\text{bH}\alpha} \geq 43.3$ . Both the weaker FUV emission and the steeper optical slope may indicate dust reddening, and their relation is explored below. Thus, we conclude that both the mean net AGN SED shape is independent of luminosity, and it also shows a small dispersion, as indicated by the distribution of  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  and  $\alpha_{\text{opt}}$  in the AGN dominated luminosity range.

### 3.7.2 The dependence of the SED on $\alpha_{\text{opt}}$

How does the overall SED change with  $\alpha_{\text{opt}}$ ? Figure 16 shows the mean SED at the eight different  $L_{\text{bH}\alpha}$  bins presented above (Fig. 11), further divided into 0.5-wide  $\alpha_{\text{opt}}$  bins. To minimize statistical noise, we present only bins with  $\geq 10$  objects. The inset in each panel shows a zoom of the spectrum around 4000Å, where the AGN emission lines of H $\delta$ , H $\epsilon$ , [Ne III], and the stellar absorption features of the Ca II H and K lines, can be seen. The stellar absorption features provide information on the host stellar population (e.g. Wild et al. 2007). Table 4 lists for each bin, the mean- $L_{\text{bH}\alpha}$ , mean- $\alpha_{\text{opt}}$ , number of objects, and detection fractions in the X-ray ( $d_X$ ) and UV ( $d_{\text{UV}}$ ). Note that at lower luminosities the mean X-ray detection fraction becomes well below 50% (Table 4), so the mean  $L_X$  becomes significantly biased. Furthermore, at  $\log L_X < 42$  the host contribution may become significant (e.g. Bauer et al. 2004).

In the highest  $\log L_{\text{bH}\alpha} = 43.5$  bin, the  $\alpha_{\text{opt}}$  distribution is narrow, as seen in Fig. 15, and there are only three  $\alpha_{\text{opt}}$  bins of SEDs. The difference between the SEDs is small in the IR, increases to the UV, and disappears again in the X-ray. This pattern is highly suggestive of a reddening effect, as the peak extinction of dust is in the UV, and it drops both towards the NIR and towards the X-ray. Also, as shown below, the H $\alpha$  EW is independent of  $\alpha_{\text{opt}}$ . If the observed SEDs were intrinsic, the large differences in the FUV emission, of more than a factor of  $> 5$ , should have produced similar amplitude in the mean H $\alpha$  EW, while the observed difference is  $\lesssim 10\%$ . These arguments support the suggestion that the red tails discussed above (Fig. 15) are due to dust extinction. It also hints that some of the dispersion in the narrow Gaussian peak of the  $\alpha_{\text{opt}}$  distribution ( $\alpha_{\text{opt}} = -0.3$  to  $-0.7$ ) is due to reddening. The reddening scenario is further explored in §3.7.3.

At  $\log L_{\text{bH}\alpha} \sim 43.0$ , there is a larger range of  $\alpha_{\text{opt}}$ , which is also consistent with a larger range of dust extinctions. However, there is a small reversal in the NIR, where the steeper  $\alpha_{\text{opt}}$  objects, which are weaker in the UV, become stronger in the NIR. The different  $\alpha_{\text{opt}}$  SEDs cross at  $\lambda \sim 1\mu\text{m}$ . One may suspect that the reverse NIR effect results from a larger covering factor of dust, and thus stronger dust emission, in more reddened objects. However, the difference is most prominent at  $\log \nu \simeq 14.5$ , where dust emission is negligible. The emission of galaxies generally peaks at  $\log \nu \simeq 14.5$ , indicating increasing host contribution as  $\alpha_{\text{opt}}$  get steeper. The lower  $L_{\text{bH}\alpha}$  bins show a stronger effect, as the NIR becomes significantly more prominent, and the SEDs crossing point shifts to shorter  $\lambda$ , going down to  $\lambda \sim 6000\text{Å}$  at  $\log L_{\text{bH}\alpha} \sim 42$ , and  $\lambda \sim 5000\text{Å}$  at  $\log L_{\text{bH}\alpha} \sim 41$ .

As the inset shows, with decreasing  $\alpha_{\text{opt}}$  at a given  $L_{\text{bH}\alpha}$ , the stellar Ca II H and K absorption lines become stronger. The effect

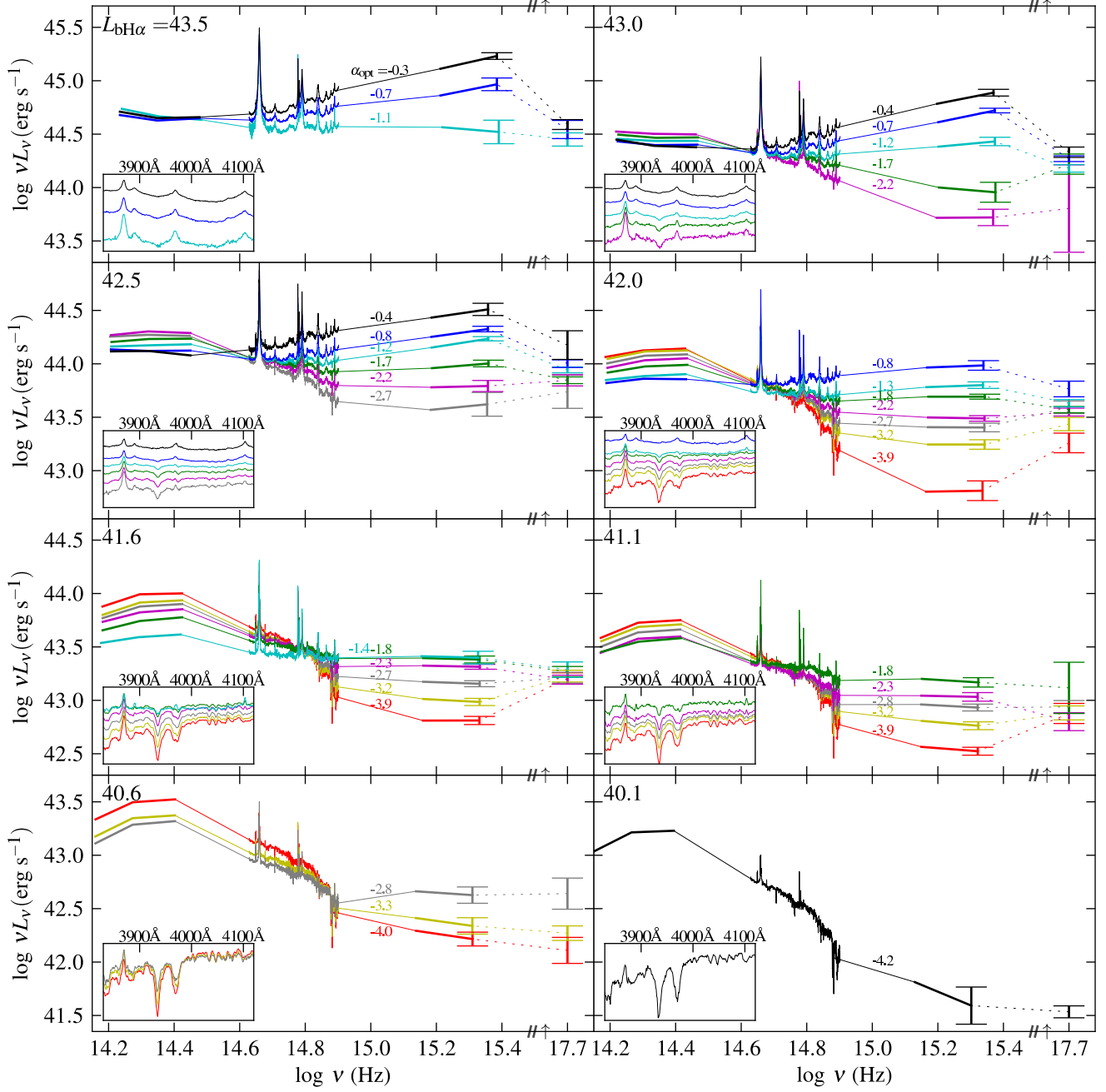
becomes more prominent at the lower  $L_{\text{bH}\alpha}$  bins. Thus, apart from the highest  $L_{\text{bH}\alpha}$  bin, the change in the SED with  $\alpha_{\text{opt}}$  also reflects an increasing host/AGN luminosity ratio, as also indicated by the increasing excess emission at  $\log \nu \simeq 14.5$ . For  $\log L_{\text{bH}\alpha} \lesssim 42$ , a steeper  $\alpha_{\text{opt}}$  at a given  $L_{\text{bH}\alpha}$ , is mostly produced by an increased host contribution, rather than by increasing dust extinction. This is indicated by the shift of the SED crossing point, i.e. spectra have a redder slope due to increasing NIR emission, rather than decreasing UV emission.

What makes some AGN be more host dominated than others, within the same  $L_{\text{bH}\alpha}$  (i.e. AGN luminosity) bin? A higher host luminosity near  $1\mu\text{m}$  implies a higher host mass, and if the host is bulge dominated, also a higher bulge mass. A higher bulge mass implies a higher  $M_{\text{BH}}$ , and thus a lower  $L/L_{\text{Edd}}$ . Thus, the range of host contributions reflects the range of  $L/L_{\text{Edd}}$  at a given AGN luminosity. This is further explored below using the relation between  $\Delta \nu$  and  $\alpha_{\text{opt}}$ , and in §3.9 by exploring directly the SED as a function of  $L/L_{\text{Edd}}$  and  $M_{\text{BH}}$ .

An association of increasing reddening with increasing host/AGN ratio is clearly seen for the  $\log L_{\text{bH}\alpha} = 42, 42.5$  AGN (Fig. 16). This suggests that the increasing reddening is associated with dust in the host galaxy. An alternative explanation for the range of reddening is dust local to the AGN, where an excess of dust along the line of sight implies a close to ‘edge on’ view, i.e. an ‘almost’ type 2 AGN. Such a local dust effect is not expected to affect host/AGN luminosity ratio near  $1\mu\text{m}$ , in contrast with the observations. We note that we cannot exclude an alternative option where the steeper  $\alpha_{\text{opt}}$  bins are composed of two separate populations, host dominated AGN, and reddened AGN, in particular if the mean broad H $\alpha$  equivalent width ( $\text{EW}_{\text{bH}\alpha}$ ) tends to be higher for host dominated AGN.

The various notions raised above are shown quantitatively in Table 4. The mean  $\text{EW}_{\text{bH}\alpha}$  of each bin is tabulated in the seventh column. In the most luminous bin, the  $\text{EW}_{\text{bH}\alpha}$  are similar to within  $\sim 10\%$ . This is consistent with the reddening scenario, as dust, which must reside outside the BLR, is unlikely to change the  $\text{EW}_{\text{bH}\alpha}$ . From the  $\log L_{\text{bH}\alpha} \sim 43$  bin and downwards, there is a general decrease of  $\text{EW}_{\text{bH}\alpha}$  with  $\alpha_{\text{opt}}$ , most likely due to the increase in host contribution. In the eighth column we list  $L_{\text{host}}/L_{\text{AGN}}(J)$ , where  $L_{\text{AGN}}(J) = 8.5 L_{\text{bH}\alpha}$  (eq. 5), and  $L_{\text{host}}(J) = \nu L_{\nu}(J) - L_{\text{AGN}}(J)$ . This column gives quantitatively the dependence of  $\alpha_{\text{opt}}$  on the relative amount of host contribution. As can be seen, the relative host contribution is a significant factor in the determination of the slope already at the second luminous bin, as concluded above based on the SED. However, the small reduction of  $\lesssim 10\%$  in the  $\text{EW}_{\text{bH}\alpha}$  with steepening  $\alpha_{\text{opt}}$ , for  $\log L_{\text{bH}\alpha} \geq 42.5$ , indicates that the steepening is not due to increasing host contribution, but due to the increased reddening. At  $\log L_{\text{bH}\alpha} \leq 42$  the significant drop in  $\text{EW}_{\text{bH}\alpha}$  with steepening  $\alpha_{\text{opt}}$  indicates the steepening is due to increasing host contribution, and not due to reddening. In column 9, the 4000Å-break is tabulated, derived from the ratio of flux at  $4000\text{Å} < \lambda < 4100\text{Å}$  to the flux at  $3850\text{Å} < \lambda < 3950\text{Å}$  ( $D_{4000}$ ; Balogh et al. 1999). The net AGN emission, and the associated ‘small blue bump’, produce  $D_{4000} < 1$ , while stars produce  $D_{4000} > 1$  which increases with decreasing surface temperature. The value of  $D_{4000}$  increases with decreasing  $L_{\text{bH}\alpha}$ , and with decreasing  $\alpha_{\text{opt}}$  at a given  $L_{\text{bH}\alpha}$ , reflecting the diminishing AGN contribution compared to old stars in the host. However, in the three lowest luminosity bins, this trend saturates for  $\alpha_{\text{opt}} < -2.5$ , and  $D_{4000}$  depends only on  $\alpha_{\text{opt}}$ , and is independent of  $L_{\text{bH}\alpha}$  (see also Fig. 16). The dependence of  $D_{4000}$  on  $\alpha_{\text{opt}}$  at these luminosities probably reflects the host type, where a bluer host galaxy (flatter





**Figure 16.** The mean IR to X-ray observed SEDs and optical spectra near 4000Å as a function of  $L_{\text{bH}\alpha}$  and  $\alpha_{\text{opt}}$ . The mean  $L_{\text{bH}\alpha}$  is noted in each panel, and the mean  $\alpha_{\text{opt}}$  is noted near each curve. Only bins with  $\geq 10$  objects are shown. The value of  $L_X$  is shifted in the y-axis (+1 dec) for presentation. The error bars denote the error in the mean  $L_X$  and  $\nu L_\nu$ (FUV). Data from different surveys are connected by thinner lines. The insets show a zoom of the optical spectrum around 4000Å, showing the AGN emission of H $\delta$   $\lambda$ 4102, He  $\lambda$ 3970, and [Ne III]  $\lambda$ 3869, and Ca II H,K  $\lambda\lambda$ 3934,3968 stellar absorption features. In the highest  $L_{\text{bH}\alpha}$  panel the difference between the SEDs increases from the NIR to the UV and disappears in the X-ray, consistent with a reddening effect. At  $42 \leq \log L_{\text{bH}\alpha} \leq 43$ , lower  $\alpha_{\text{opt}}$  SEDs also show a stronger bump at  $\log \nu = 14.4$ , and stronger stellar absorption features, indicating a stronger host contribution, in addition to reddening. At  $L_{\text{bH}\alpha} \leq 41$ , the host dominates the emission at all bands (see Figs. 10, 13), and the increase of the equivalent widths of the stellar absorption features with decreasing  $\alpha_{\text{opt}}$  reflects the change in the host stellar population. Note that dust extinction appears as missing UV emission for a steeper  $\alpha_{\text{opt}}$ , while host contribution appears as excess NIR emission for a steeper  $\alpha_{\text{opt}}$ . The host contribution progresses to shorter wavelengths with decreasing  $L_{\text{bH}\alpha}$ .

$L_{\text{bH}\alpha}$ range (log erg s $^{-1}$ )	$L_{\text{bH}\alpha}$ (log erg s $^{-1}$ )	$\alpha_{\text{opt}}$	$N$	detection fractions $d_{\text{UV}}$	$d_X$	$\text{EW}_{\text{bH}\alpha}$ (Å)	$\frac{L_{\text{host}}}{L_{\text{AGN}}}$ (J)	$D_{4000}$	$\Delta v$ (km s $^{-1}$ )
43.3 - 43.8	43.50	-0.30	66	1.00	0.73	453 ( $\pm 13$ )	0.8	0.92	4000
	43.45	-0.66	38	1.00	0.71	456 ( $\pm 20$ )	1.0	0.93	4000
	43.42	-1.13	11	0.89	0.55	503 ( $\pm 34$ )	1.0	0.94	4500
42.8 - 43.3	43.07	-0.35	64	1.00	0.89	353 ( $\pm 13$ )	1.6	0.93	2900
	43.03	-0.74	145	0.97	0.66	349 ( $\pm 9$ )	2.1	0.94	3500
	42.98	-1.20	102	1.00	0.59	318 ( $\pm 10$ )	2.7	0.97	4100
	42.98	-1.70	32	0.96	0.41	296 ( $\pm 17$ )	3.0	1.02	5000
	42.97	-2.23	10	0.90	0.20	303 ( $\pm 27$ )	3.2	1.07	7000
42.3 - 42.8	42.62	-0.39	14	1.00	0.71	217 ( $\pm 24$ )	3.2	0.95	2600
	42.60	-0.80	127	0.97	0.68	252 ( $\pm 8$ )	3.4	0.96	2700
	42.55	-1.24	211	0.97	0.58	220 ( $\pm 6$ )	4.7	1.00	3300
	42.52	-1.71	153	1.00	0.49	198 ( $\pm 6$ )	5.9	1.04	4000
	42.49	-2.22	75	0.91	0.56	189 ( $\pm 9$ )	7.3	1.09	4300
	42.48	-2.72	25	0.64	0.36	200 ( $\pm 15$ )	6.7	1.14	4900
41.8 - 42.3	42.13	-0.81	40	0.97	0.80	145 ( $\pm 10$ )	6.5	0.98	2300
	42.08	-1.29	126	0.95	0.60	153 ( $\pm 6$ )	7.8	1.02	2300
	42.06	-1.76	226	0.99	0.50	125 ( $\pm 3$ )	10.3	1.08	2700
	42.02	-2.24	217	0.95	0.40	112 ( $\pm 3$ )	13.3	1.13	3400
	42.01	-2.70	122	0.93	0.38	107 ( $\pm 4$ )	14.7	1.20	3800
	41.98	-3.21	62	0.85	0.29	93 ( $\pm 6$ )	18.0	1.31	4200
	41.97	-3.86	29	0.70	0.41	94 ( $\pm 9$ )	19.1	1.47	5400
41.3 - 41.8	41.64	-1.35	28	0.96	0.68	108 ( $\pm 9$ )	12.1	1.04	1900
	41.61	-1.79	115	0.97	0.36	78 ( $\pm 4$ )	20.1	1.10	2200
	41.56	-2.27	212	0.96	0.35	64 ( $\pm 2$ )	25.2	1.18	2600
	41.56	-2.74	186	0.93	0.31	61 ( $\pm 2$ )	28.5	1.25	3100
	41.54	-3.25	146	0.88	0.31	56 ( $\pm 2$ )	31.5	1.36	3900
	41.52	-3.86	129	0.78	0.22	47 ( $\pm 2$ )	38.5	1.51	4500
40.8 - 41.3	41.17	-1.76	20	0.95	0.30	45 ( $\pm 5$ )	35.0	1.15	1800
	41.11	-2.31	59	0.96	0.32	41 ( $\pm 3$ )	41.3	1.21	2400
	41.10	-2.77	127	0.96	0.30	38 ( $\pm 2$ )	49.2	1.29	2700
	41.08	-3.24	111	0.91	0.18	33 ( $\pm 1$ )	55.4	1.41	3000
	41.06	-3.91	160	0.81	0.12	29 ( $\pm 1$ )	63.2	1.60	4100
40.3 - 40.8	40.62	-2.80	23	0.95	0.09	32 ( $\pm 4$ )	64.9	1.29	2500
	40.62	-3.26	30	0.90	0.27	28 ( $\pm 3$ )	76.3	1.42	3100
	40.59	-3.98	74	0.85	0.20	20 ( $\pm 1$ )	110.8	1.63	3900
39.8 - 40.3	40.10	-4.24	10	0.75	0.30	15 ( $\pm 2$ )	182.0	1.66	4100

**Table 4.** Mean quantities of the bins of  $L_{\text{bH}\alpha}$  and  $\alpha_{\text{opt}}$  pictured in Fig. 16. The  $L_{\text{bH}\alpha}$  ranges are as in Fig. 11 and Table 2. The  $\alpha_{\text{opt}}$  bins have a width of 0.5. Only bins with  $N \geq 10$  are listed. Col. 6 shows the mean and the error in the mean  $\text{EW}_{\text{bH}\alpha}$ . The  $L_{\text{host}}/L_{\text{AGN}}$  ratio (Col. 7), is calculated as in Fig. 13. The formulation of the 4000Å-break (Col. 8) is given in Balogh et al. (1999).

$\alpha_{\text{opt}}$ ) is associated with hotter stars, and thus lower  $D_{4000}$  values. Finally, the last column shows the average  $\Delta v$  of each bin. The  $\Delta v$  follows  $\alpha_{\text{opt}}$  at all  $L_{\text{bH}\alpha}$ , similar to  $D_{4000}$  and  $L_{\text{host}}/L_{\text{AGN}}$ (J). This trend reflects the  $M_{\text{BH}}$  versus bulge mass relation. As discussed above, a steeper  $\alpha_{\text{opt}}$  is associated with a more massive bulge, at a given AGN luminosity (given  $L_{\text{bH}\alpha}$  bin). A higher bulge mass implies a higher  $M_{\text{BH}}$ , and therefore necessarily requires a higher  $\Delta v$ . Since  $M_{\text{BH}} \propto \Delta v^2$  (eq. 2), at a fixed AGN luminosity, we expect a steeper than linear rise in  $L_{\text{host}}/L_{\text{AGN}}$ (J) with  $\Delta v$ , as Table 4 indicates. Also, at the three lowest  $L_{\text{bH}\alpha}$  bins, which are dominated by the host emission, a bluer host colour is associated with a lower  $\Delta v$ , indicating a lower bulge mass, and thus a later type host galaxy. This result is consistent with the tendency of bluer galaxies to be of later type.

### 3.7.3 The effect of dust extinction on the SED

Above we have shown that the effect of the host on the optical slope is small at  $L_{\text{bH}\alpha} \sim 10^{43.5}$  erg s $^{-1}$ , and that the dispersion in optical slope at this luminosity seems to originate from a reddening effect. In this subsection we evaluate the implied general properties of the extinction law.

The upper panel of Figure 17 compares  $vL_V(\text{UV})/L_{\text{bH}\alpha}$  with  $\alpha_{\text{opt}}$  in the T1 objects from the  $L_{\text{bH}\alpha} = 43.5$  bin. The  $vL_V(\text{UV})$  value is the geometrical mean of the NUV and FUV luminosities. There is a strong correlation of  $\alpha_{\text{opt}}$  versus  $vL_V(\text{UV})/L_{\text{bH}\alpha}$  ( $P_r = 0.78$ , significance of  $11\sigma$ ), with a best-fitting least squares slope of  $0.93 \pm 0.06$ . This correlation strongly suggests reddening is a significant cause of dispersion in  $\alpha_{\text{opt}}$ , already in the  $-1 < \alpha_{\text{opt}} < 0$  range ( $P_r = 0.5$ , significance of  $6\sigma$  for objects with  $\alpha_{\text{opt}} > -1$ ).

Over plotted on the data are the expected relations for different extinction curves, calculated using the formulations in Pei



(1992) for MW, SMC and LMC dust, and also the ‘grey dust’ formulation in Gaskell & Benker (2007). We assume the intrinsic unabsorbed mean values of  $\alpha_{\text{opt}} = -0.3$  which extends to the UV, and  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha} = 47$ . We apply the appropriate extinction law, and then convolve the extinguished power law with the response functions of the GALEX filters at  $z = 0.24$  (the mean  $z$  of the  $\log L_{\text{bH}\alpha} = 43.5$  bin, Table 2). The SMC or LMC extinction laws are preferred over MW and grey dust.

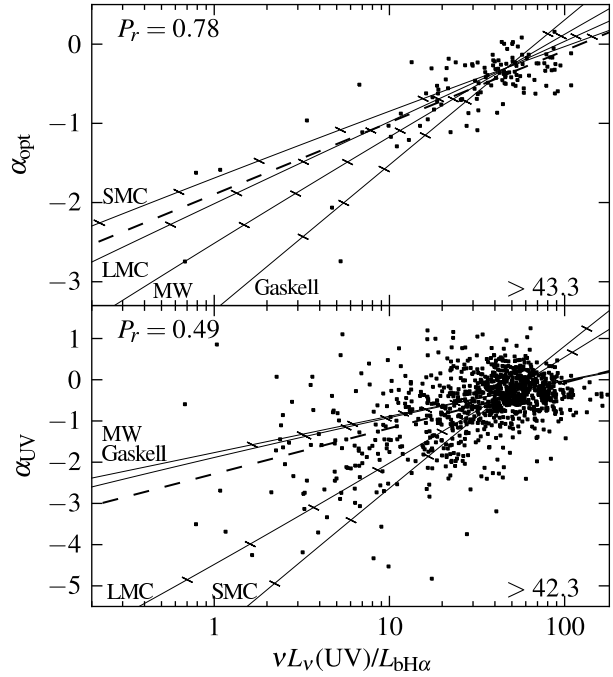
An SMC extinction correction of  $\Delta\alpha_{\text{opt}} = 0.4$  translates to  $E(B-V) = 0.1$ . Therefore, the  $\alpha_{\text{opt}} = -0.7$  bin (30% of the objects, Table 4) and  $\alpha_{\text{opt}} = -1.1$  bin (9% of the objects) have an  $E(B-V)$  of 0.1 and 0.2 larger than the  $\alpha_{\text{opt}} = -0.3$  bin. The mean  $E(B-V)$  of the  $\alpha_{\text{opt}} = -0.3$  bin, which includes the majority of the objects, is hard to assess, as there may be some intrinsic dispersion in  $\alpha_{\text{opt}}$  (the best fitting slope in Fig. 17 yields a residual scatter of 0.29 in  $\alpha_{\text{opt}}$ ). If the 10 percentile of flattest  $\alpha_{\text{opt}}$  of all  $\log L_{\text{bH}\alpha} = 43.5$  T1 objects,  $-0.13$ , is the intrinsic unreddened slope, then the mean  $E(B-V)$  of the  $\alpha_{\text{opt}} = -0.3$  bin is  $\sim 0.04$ .

In the lower panel we compare the UV slope  $\alpha_{\text{UV}} (\equiv 5.8 \times \log(l_{\nu;\text{FUV}}/l_{\nu;\text{NUV}}))$  with  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$  for the T1 objects with  $L_{\text{bH}\alpha} > 10^{42.3} \text{ erg s}^{-1}$ , where the host emission in the NUV is negligible (note  $\alpha_{\text{UV}}$  and  $\nu L_{\nu}(\text{UV})$  are independent quantities). The general trend is a decreasing  $\alpha_{\text{UV}}$  with decreasing  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$ , as expected from reddening, but the scatter is significantly larger than the relation above with  $\alpha_{\text{opt}}$ . The mean trend seems to favor the MW or the grey dust curve. Note that unlike  $\alpha_{\text{opt}}$ , which is based on the continuum derived from the SDSS spectroscopy,  $\alpha_{\text{UV}}$  and  $\nu L_{\nu}(\text{UV})$  are derived from broad band photometry, which includes the UV continuum and emission features. The dispersion in emission line properties likely induces the large scatter observed. We also note that the extinction law lines are derived based on the mean  $z$  of 0.18, while the apparent amount of reddening changes with  $z$  as different rest-frame frequencies enter the filters. This is especially pronounced in the MW extinction curve, as the 2175Å feature enters and leaves the two UV filters. UV spectroscopy is required to derive more accurate continuum slopes in the UV, and to test whether the extinctions laws implied by the UV and by the optical are indeed inconsistent.

### 3.8 The dependence of the SED on $L_{\text{bH}\alpha}$ and $\Delta v$

In the following section we briefly explore the dependence of the mean SED on  $L/L_{\text{Edd}}$  and  $M_{\text{BH}}$ . Since both parameters are derived from the two directly observed quantities,  $L_{\text{bH}\alpha}$  and  $\Delta v$ , we first explore the dependence of the mean SED on these two quantities.

Figure 18 presents the mean optical spectra as a function of  $L_{\text{bH}\alpha}$  and  $\Delta v$ . Each panel is for one decade wide  $L_{\text{bH}\alpha}$  bins, and the plots in each panel are for  $\log \Delta v$  bins a quarter decade wide. Various prominent emission lines and Fe II multiplets are noted in the plot (Vanden Berk et al. 2001; Phillips 1978). The inset in the lower two panels zooms on the spectrum near 4000Å (see Fig. 16). The highest luminosity AGN at  $\log L_{\text{bH}\alpha} = 43.1$  (upper panel) show clearly the strong dependence of the Fe II emission on  $\Delta v$ , which is part of the Boroson & Green (1992) eigenvector 1 (EV1) trends. The  $\log \Delta v = 3.2$  spectrum shows that Fe II multiplets are present from 7000Å to below 4200Å. The Balmer lines become asymmetric, with increasing  $\Delta v$ , as expected from the EV1 correlations. The EW of the [O I]  $\lambda 6300$ , [O II]  $\lambda 3727$ , and [Ne III]  $\lambda 3869$  also increase with increasing  $\Delta v$ . Stellar absorption features such as Ca II H,K  $\lambda \lambda 3934, 3968$ , G4300, Mg b  $\lambda 5175$ , and NaD  $\lambda 5893$  become clear at the highest  $\log \Delta v = 4.1$  plot, and some can be traced back to lower  $\Delta v$  values.

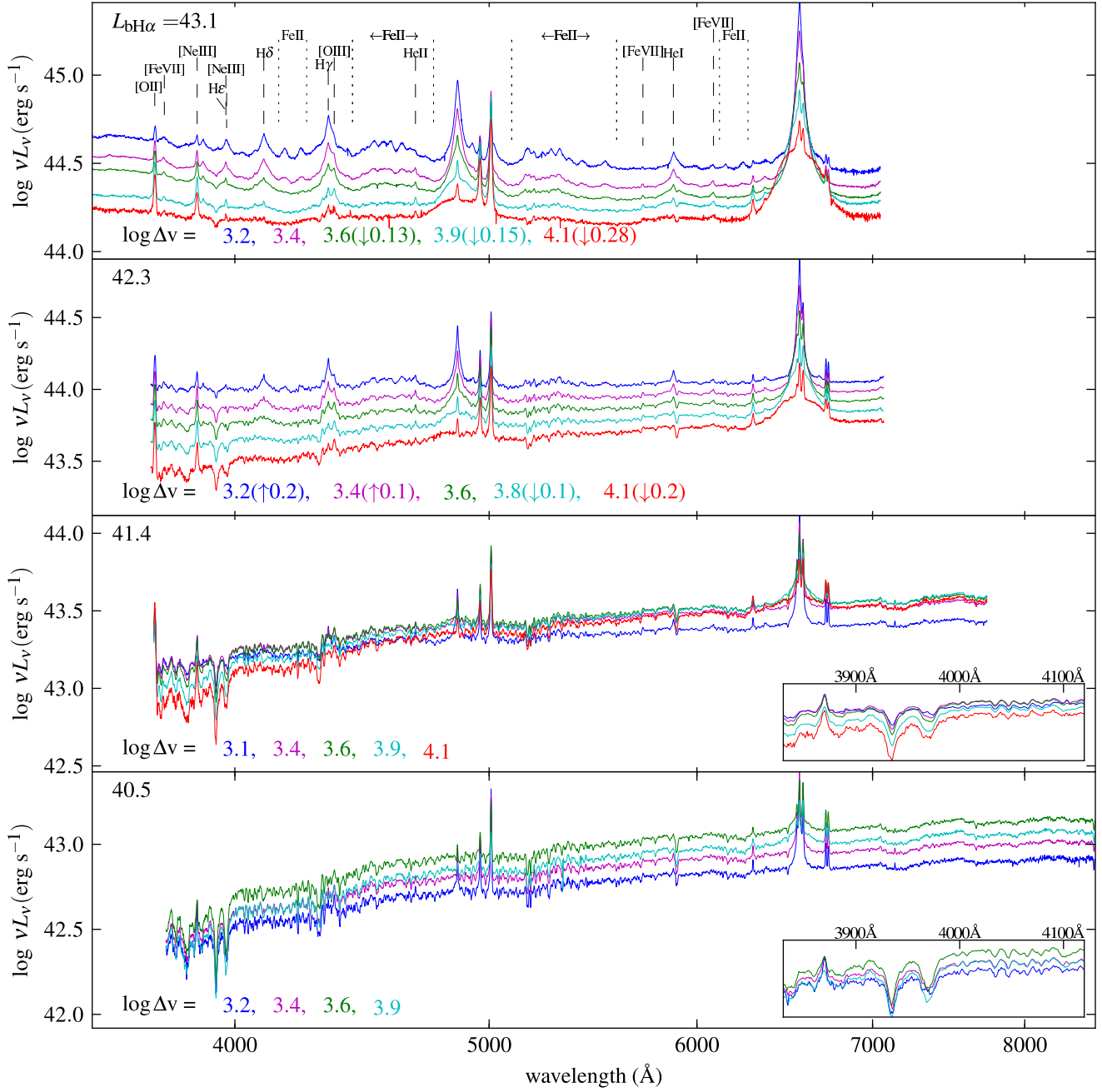


**Figure 17.** The observed relation of  $\alpha_{\text{opt}}$  and  $\alpha_{\text{UV}}$  with  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$ . Over plotted are the expected relations for different extinction laws of an intrinsic power law emission with  $\alpha_{\text{opt}} = \alpha_{\text{UV}} = -0.3$  and a  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha} = 47$ . Small dashes ‘-’ mark jumps of 0.1 in  $E(B-V)$ . The minimum  $\log L_{\text{bH}\alpha}$  of the T1 objects used, to avoid host contribution, are noted in each panel. The data points are fit by linear relations (dashed lines). **Upper panel** The two independent ratios are highly correlated, and follow an SMC/LMC extinction law. This correlation suggests that reddening affects the dispersion in  $\alpha_{\text{opt}}$  also for  $\alpha_{\text{opt}} > -1$  (the narrow Gaussian peak of the  $\alpha_{\text{opt}}$  and  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$  distributions, Fig. 15). **Lower panel** There is a general decrease of  $\alpha_{\text{UV}}$  with  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$ , consistent with reddening, though with substantial scatter. The mean trend prefers the MW / Gaskell & Benker (2007) extinction curves, although photometry-based slopes are significantly uncertain.

The  $\log L_{\text{bH}\alpha} = 42.3$  bin shows a more prominent host contribution with increasing  $\Delta v$ , as indicated by the increasing stellar absorption features and by the steepening  $\alpha_{\text{opt}}$ . These effects are more prominent than in the upper panel. However the increase in the narrow line EW with increasing  $\Delta v$  is not as strong as in the  $\log L_{\text{bH}\alpha} = 43.1$  bin. The lowest two panels show spectra which are dominated by the host continuum. This is demonstrated by the fact that the mean continuum luminosity density in the  $\log L_{\text{bH}\alpha} = 41.4$  bin is typically only a factor of two lower than in the  $\log L_{\text{bH}\alpha} = 42.3$  panel, although the expected net AGN contribution, based on  $L_{\text{bH}\alpha}$ , drops by an order of magnitude.

### 3.9 The dependence of the SED on $L/L_{\text{Edd}}$ and $M_{\text{BH}}$

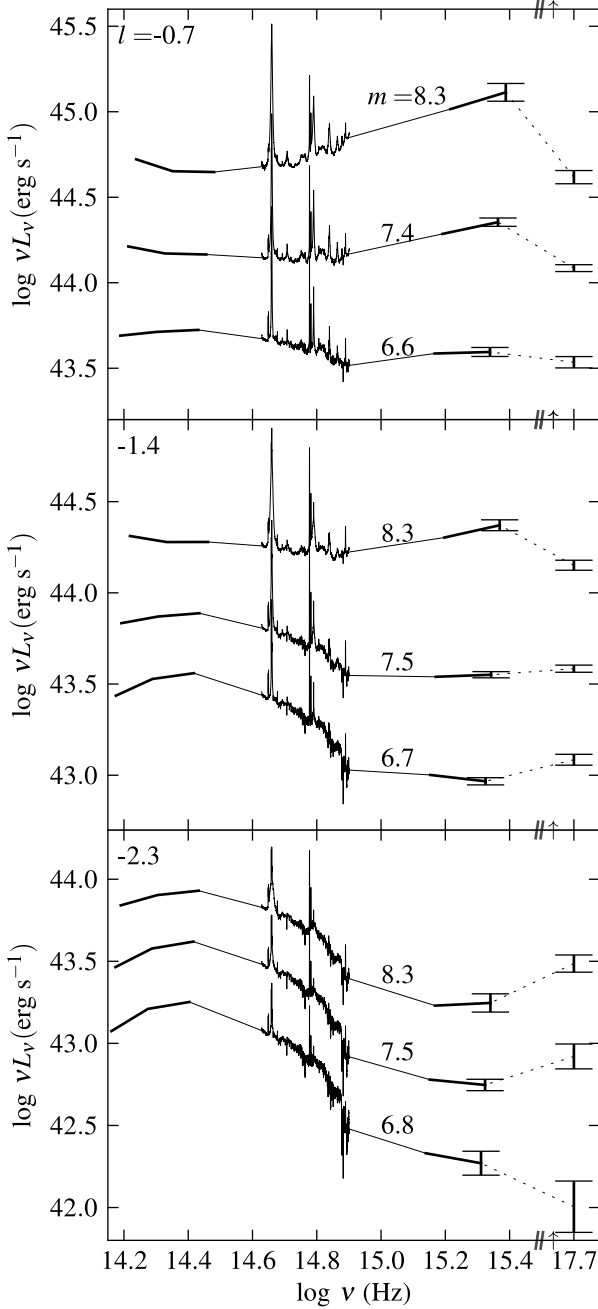
Figure 19, presents the SED binned by  $L/L_{\text{Edd}}$ , for each panel, and by  $M_{\text{BH}}$ , within each panel. The bins are a decade wide, with  $L/L_{\text{Edd}}$  in the range  $-3 < l < 0$  and  $M_{\text{BH}}$  in the range  $6 < m < 9$ . With decreasing  $L/L_{\text{Edd}}$  the host contribution becomes more dominant, as expected since  $L/L_{\text{Edd}}$  provides a measure of  $L/L_{\text{bulge}}$ .



**Figure 18.** The mean optical spectra as a function of  $L_{bH\alpha}$  and  $\Delta v$  (listed in each panel). To increase clarity, some mean spectra were shifted, as noted in the legends. Some prominent emission lines and Fe II multiplets are marked. The inset zooms on the spectrum near 4000 Å, as in Fig. 16. The highest  $L_{bH\alpha}$  bin shows the increasing EV1 emission (BG92) with decreasing  $\Delta v$ . The BG92 analysis was near H $\beta$  and this plot shows the extension of the EV1 features from 7000 Å to below 4000 Å. Note the change in the Fe II multiplets near 4250 Å and 6250 Å, and in the various forbidden lines at  $\lambda > 5500$  Å and  $\lambda < 4500$  Å. At  $\log L_{bH\alpha} = 42.3$  there is an increase in stellar absorption features with increasing  $\Delta v$ , in addition to the EV1 trend. Higher  $\Delta v$  implies a higher  $M_{BH}$  and thus a more massive and more luminous bulge (up to the SDSS aperture effect) for the same mean AGN luminosity. At  $\log L_{bH\alpha} \leq 41.4$ , the host dominates the continuum, and at  $\log L_{bH\alpha} = 40.5$  the broad H $\beta$  becomes undetectable due to the host contribution. Such objects will be defined as net type 2 AGN based on H $\beta$  only.

However, we also see a trend where the host contribution becomes more dominant (the NIR region of the SED turning from a dip into a peak) with decreasing  $M_{BH}$  at a given  $L/L_{Edd}$ , in particular for  $l \geq -1.4$  (the two upper panels). This likely arises from the fact that  $L_{host}/L_{bulge}$  rises with decreasing  $L_{bulge}$ , as expected in a flux

limited sample (SDSS), where the low luminosity bulges are more commonly found in more luminous disc galaxies, rather than in lower luminosity bulge-only galaxies (i.e. dwarf ellipticals), which will be below the detection limit. Thus, Fig. 19 indicates that at a fixed  $l$ , the host tends to become disc dominated with decreasing  $m$ .



**Figure 19.** The mean observed SED as a function of  $L/L_{\text{Edd}}$  and  $M_{\text{BH}}$ . The bins are a decade wide in both  $m$  and  $l$ , in the ranges  $-3 < l < 0$  and  $6 < m < 9$ . Each panel is for a given  $l$  bin (mean values in the upper left corner), subdivided by  $m$  (mean values near each curve). The values of  $L_X$  are shifted in the y-axis (+1 dec) for presentation. The error bars denote the error in the mean  $L_X$  and  $\nu L_{\nu}(\text{FUV})$ . Note the increase in the relative host contribution (the bump near  $\log \nu = 14.4$ ) with decreasing  $l$  (top to bottom panel), as expected since  $L_{\text{AGN}}/M_{\text{BH}}$  corresponds to  $L_{\text{AGN}}/L_{\text{bulge}}$  and the T1 sample is bulge dominated (§3.5). Note also the increase in the relative host contribution with decreasing  $m$ , at a given  $l$ , in the middle and upper panels. This likely reflects the increase in  $L_{\text{host}}/L_{\text{bulge}}$  with decreasing  $m$ , as smaller bulges are often found in disc galaxies. At  $l < -2$  (lowest panel), the optical-NUV SED does not change with  $M_{\text{BH}}$ , as the host dominates for all  $M_{\text{BH}}$ . The fraction of X-ray detections is  $> 50\%$  only in the four top curves, so other plotted  $L_X$  values can be significantly biased.

In the lowest  $L/L_{\text{bulge}}$  bin ( $l = -2.3$ ) the host dominates at all  $m$ , although some AGN contribution can be noticed in the FUV and X-ray, in particular at  $m = 8.3$ , which corresponds to the highest AGN luminosity.

The SED trends seen in Fig. 19 are dominated by the host contribution. To explore trends in the net AGN SED with  $L/L_{\text{Edd}}$  and  $M_{\text{BH}}$ , we use the emission bands least affected by the host contribution, which are the FUV and X-ray bands (e.g. Figs. 10, 13). Following the discussions in §3.3 and §3.6, in order to avoid host contribution in these bands, we utilize only objects with  $\log L_{\text{bH}\alpha} > 42$  when analyzing the FUV emission, and objects with  $\log L_{\text{bH}\alpha} > 41$  when analyzing X-ray emission. Also, to reduce detection biases when analyzing X-ray emission, we use only objects with  $F_{\text{bH}\alpha} > 10^{-13.5} \text{ erg s}^{-1} \text{ cm}^{-2}$  (Fig. 3, §3.6), where the X-ray detection rate is  $> 50\%$ .

Below we explore which of the parameters,  $L_{\text{bH}\alpha}$ ,  $\Delta\nu$ ,  $M_{\text{BH}}$  and  $L/L_{\text{Edd}}$ , appears to be the dominant parameter which drives the values of three SED parameters,  $L_X/L_{\text{bH}\alpha}$ ,  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$ , and their ratio  $\alpha_{\text{FUV},x} \equiv 0.42 \times \log l_{\nu}(2\text{keV})/l_{\nu}(\text{FUV})$ . Since  $L_{\text{bH}\alpha} \propto L_{\text{bol}}$  (eq. 6), the ratios of  $L_X$  and of  $\nu L_{\nu}(\text{FUV})$  to  $L_{\text{bH}\alpha}$  is a measure of the fraction of the bolometric luminosity emitted at each band. The value of  $\alpha_{\text{FUV},x}$  is closely related to the more commonly used 2500Å–2 keV slope ( $\alpha_{\text{ox}}$ ), which in eq. 8 was calculated by deriving the 2500Å luminosity from  $L_{\text{bH}\alpha}$  (eq. 5).

The binning into  $L/L_{\text{Edd}}$  and  $M_{\text{BH}}$  is performed in the following manner. The objects are sorted by either  $m$  or by  $l$  and divided into four equal size groups. Each of these groups is then sorted by the other property, and again divided into four equal size groups. This ensures similar statistical errors in all bins. We repeat this processes by sorting based on  $\Delta\nu$  and  $L_{\text{bH}\alpha}$ .

Figure 20, left column, presents the derived relations of the mean values of the three SED parameters as a function of  $l$ , at a fixed  $m$ , for  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$ , and as a function of  $m$  at a fixed  $l$  for  $L_X/L_{\text{bH}\alpha}$  and for  $\alpha_{\text{FUV},x}$ . Error bars denote the error in the mean. The upper panels show that  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  is largely set by  $l$ , and is independent of  $m$ . The drop of  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  steepens with  $l$ , with a drop mostly at  $l < -1.3$ . The right hand column of Fig. 20 presents  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  as a function of  $\Delta\nu$  for different bins in  $L_{\text{bH}\alpha}$ . Both parameters contribute, as  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  decreases with increasing  $\Delta\nu$ , at a fixed  $L_{\text{bH}\alpha}$ , and also with decreasing  $L_{\text{bH}\alpha}$ , at a fixed  $\Delta\nu$ . The comparison of the upper two panels indicates that  $l$  is the main driver of the  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  distribution, and the trends with  $\Delta\nu$  and  $L_{\text{bH}\alpha}$  are driven by the  $l$  dependence.

As discussed above, the tail towards lower  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  values (Figs. 15, 17) is likely due to reddening. The mean  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha} = 18$  at  $l = -2$  (Fig. 20) may result from an increased amount of reddening with decreasing  $L/L_{\text{Edd}}$ , in particular close to  $L/L_{\text{Edd}} = -2$ . A trend of  $\nu L_{\nu}(\text{FUV})/L_{\text{bH}\alpha}$  with  $l$  is consistent with the trend described above of an increasing reddening with the host/AGN luminosity ratio.

Figure 20, middle row panels, explore the driving parameter of  $L_X/L_{\text{bH}\alpha}$ . Here there are mixed results. There is a general trend of decreasing  $L_X/L_{\text{bH}\alpha}$  with increasing  $m$ , but with some  $l$  dependence of the amplitude for  $m > 7.5$ , where a higher  $l$  implies a lower  $L_X/L_{\text{bH}\alpha}$ . There is also a general trend of decreasing  $L_X/L_{\text{bH}\alpha}$  with increasing  $L_{\text{bH}\alpha}$ , with a weaker trend of decreasing  $L_X/L_{\text{bH}\alpha}$  with increasing  $\Delta\nu$ . None of the four parameters appear to be a dominant parameter.

Figure 20, lower row panels, explore the driving parameter of  $\alpha_{\text{FUV},x}$ . The slope gets steeper with increasing  $m$  and with increasing  $l$ , and both parameters contribute similarly to the trends. The right hand panel shows that  $\alpha_{\text{FUV},x}$  is mostly driven by  $L_{\text{bH}\alpha}$ . There

is only a weak dependence of  $\alpha_{\text{FUV},x}$  on  $\Delta v$ . This indicates that the dependence of  $\alpha_{\text{FUV},x}$  on  $M_{\text{BH}}$  and  $L/L_{\text{Edd}}$  found by Kelly et al. (2008) is probably due to the correlation between these parameters and AGN luminosity.

## 4 DISCUSSION

### 4.1 The distribution of $\Delta v$

Croom (2011) already noted that the mean value of  $\Delta v$  of the H $\beta$  line in the SDSS quasar samples is independent of luminosity. Here we extend this result and show that the distribution of  $\Delta v$  values is also independent of luminosity (Fig. 7). Why is this result surprising? The observed value of  $\Delta v$  is determined by the AGN luminosity, which sets the radius of the BLR, and by  $M_{\text{BH}}$ , which sets the velocity dispersion at that radius. Thus, the distribution of  $\Delta v$  value,  $f(\Delta v)$ , is set by the AGN black hole mass function,  $f(M_{\text{BH}})$ , and by the AGN luminosity function,  $f(L_{\text{bol}})$ , which may vary with  $M_{\text{BH}}$ . The observed  $f(\Delta v)$  is also set by the detection efficiency of individual broad line AGN. As shown above, selection effects are not expected to dominate  $f(\Delta v)$  (Fig. 6, upper panel), thus the lack of dependence of  $f(\Delta v)$  on  $L_{\text{bH}\alpha}$  appears to be an intrinsic property of the SDSS T1 sample. At face value, the physical implications are that the mean  $M_{\text{BH}}$  increases roughly as  $L_{\text{bol}}^{1/2}$ , and similarly the mean  $L/L_{\text{Edd}}$  also scales as  $L_{\text{bol}}^{1/2}$  (Fig. 6 lower panel). These relations reflect the simple fact that the distribution of the objects along the  $L_{\text{bol}}$  axis and along the  $\Delta v$  axis are independent of each other. The only significant deviations are the rise in the lower limit of  $\Delta v$  with  $L_{\text{bol}}^{1/4}$ , consistent with the physical Eddington limit, i.e.  $l = 0$ , and the drop in the upper limits on  $\Delta v$  with  $L_{\text{bol}}^{-1/4}$ , consistent with an upper limit to  $M_{\text{BH}}$  of a few  $10^9 M_{\odot}$ . The convergence of these upper and lower limits on  $\Delta v$  at  $L_{\text{bol}} > \text{a few } 10^{45} \text{ erg s}^{-1}$  leads to a narrower  $f(\Delta v)$ , which produces a small bump in  $f(\Delta v)$  at  $\Delta v_{1000} \sim 7$  (Fig. 7). A similar convergence of the range of FWHM for the Mg II and C IV lines at the highest continuum luminosities was noted by Fine et al. (2008, 2010).

Could there be another physical effect which controls the form of  $f(\Delta v)$ , other than the black hole mass function and the AGN luminosity function? Laor & Davis (2011) noted that the accretion disc continuum becomes colder with increasing  $\Delta v$ , and for a large enough value (depending on the black hole spin), the accretion disc can become non-ionizing, and thus form a lineless quasar which is harder to detect. This may explain the rarity of objects with  $\Delta v > 10,000 - 20,000 \text{ km s}^{-1}$ . However, it cannot explain the similarity of  $f(\Delta v)$  at  $\Delta v < 10,000 \text{ km s}^{-1}$ , which therefore remains an open question.

### 4.2 The SED

#### 4.2.1 A uniform blue optical-UV SED

As described above, at  $L_{\text{bH}\alpha} > 10^{42} \text{ erg s}^{-1}$  the net AGN SED in the 9000Å–1500Å range is consistent with a fixed shape, and an amplitude which scales linearly with  $L_{\text{bH}\alpha}$ . The SED shape is consistent with the R06 SED of luminous AGN. Furthermore, the linearity of the mean  $L_{\text{bH}\alpha}$  with  $\nu L_{\nu}$  in the optical-UV range suggests that the fixed shape SED extends to the ionizing continuum as well. The linearity breaks in the X-ray regime, where the relative X-ray emission drops with increasing luminosity (Fig. 20, and various earlier studies, e.g. Just et al. 2007). Other studies (Maiolino

et al. 2007; Treister et al. 2008) indicate there is also a non linearity in the mid-IR emission as well, which also decreases in relative strength with increasing luminosity.

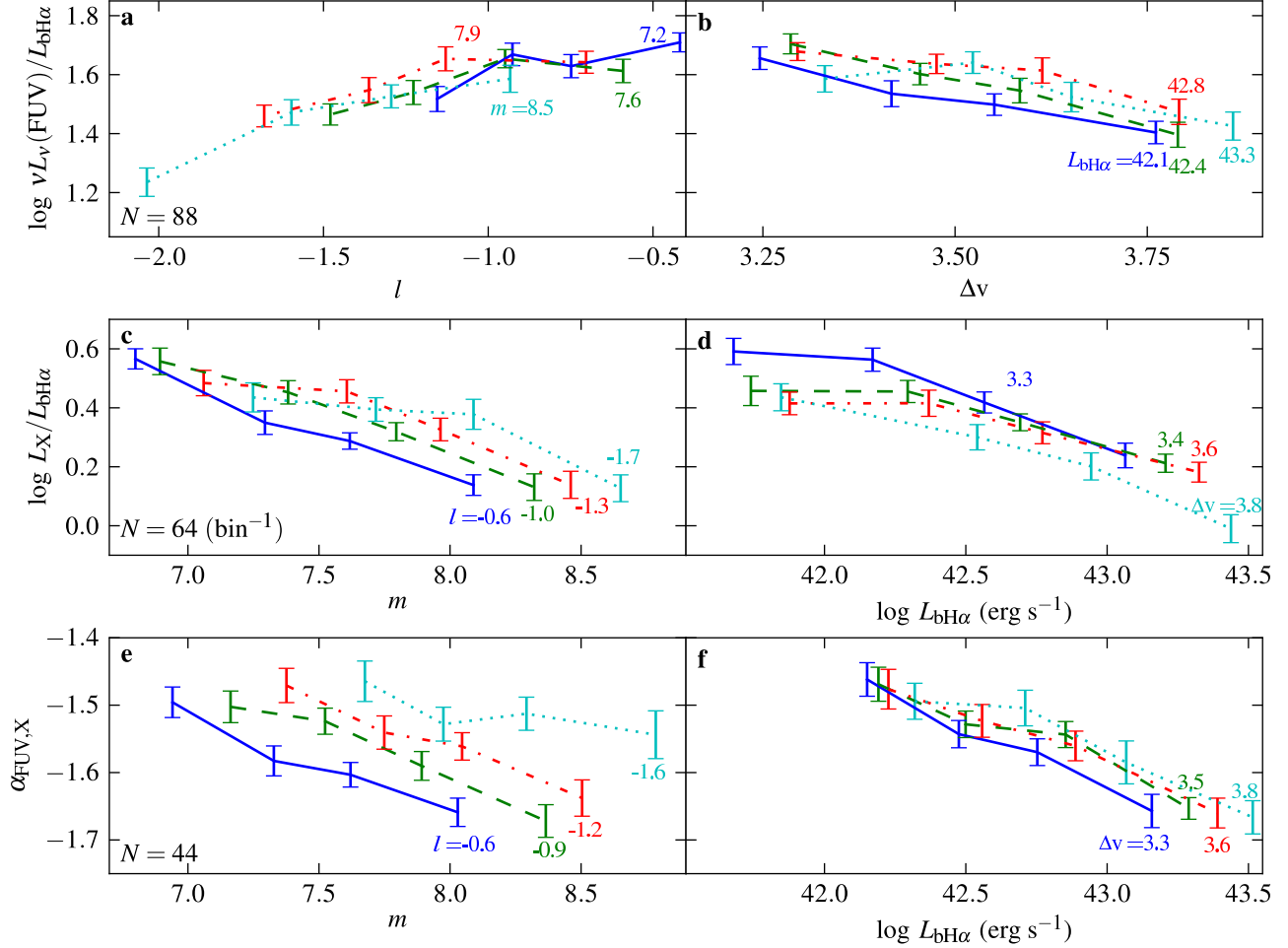
The dispersion in the SED shapes, as indicated by the observed distribution of  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$  and  $\alpha_{\text{opt}}$  values (Fig. 15), is driven at lower luminosities by the range of host contributions to the SED. The luminous AGN ( $\log L_{\text{bH}\alpha} = 43.5$  bin) show a significantly smaller dispersion. Furthermore, for  $-1 < \alpha_{\text{opt}} < 0$ , within the narrow Gaussian peak of the distribution, there is a correlation of  $\alpha_{\text{opt}}$  with  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$ , consistent with reddening (Fig. 17). This suggests that the intrinsic dispersion in  $\alpha_{\text{opt}}$ , after correction for reddening, is smaller than observed. These results are consistent with earlier suggestions (Ward et al. 1987; Gaskell et al. 2004) that AGN have a single shape SED over a large luminosity range, which is affected by dust and host contribution.

A rather uniform and blue intrinsic optical-UV SED is expected if AGN are powered by a thin accretion disc, which emits locally close to a blackbody. A locally blackbody Newtonian accretion disc produces the well known  $L_{\nu} \propto \nu^{1/3}$  SED at wavelengths emitted from intermediate regions in the disc, away from the inner and outer edges of the disc. Including non-LTE effects in the disc atmospheric emission, and the relativistic corrections for the disc structure and for the observed emission, leads to an optical-UV slopes somewhat steeper than  $\nu^{1/3}$ , but still bluer than the average observed slope (e.g. Bonning et al. 2007). Davis et al. (2007) found that a small amount of reddening of  $E(B - V) \sim 0.03 - 0.055$  of SMC-like dust, is sufficient to make the observed UV SED consistent with the accretion disc model predictions. These values are consistent with a mean  $E(B - V) \sim 0.04$  for SMC-like dust, deduced above (§3.7.3) by comparing the slope of the 10 percentile of flattest luminous AGN  $\alpha_{\text{opt}} = -0.13$ , with the peak of the narrow distribution at  $\alpha_{\text{opt}} = -0.3$ , where 60% of the AGN reside (Table 4).

#### 4.2.2 The reddening dust

What is the fraction of AGN affected by reddening? As noted above, in the most luminous AGN bin ( $\log L_{\text{bH}\alpha} = 43.5$ ), about 60% of the AGN may be slightly affected by reddening with  $E(B - V) \sim 0.04$ , as indicated by the correlation of  $\alpha_{\text{opt}}$  vs.  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$ , consistent with dust extinction affecting the SED at  $-1 < \alpha_{\text{opt}} < 0$  (Fig. 17). An additional 30% of these AGN, have a mean  $\alpha_{\text{opt}} = -0.7$  (Table 4), which may correspond to a reddening of  $E(B - V) \sim 0.1$  (§3.7.3), and the remaining 10% which reside in the  $\alpha_{\text{opt}} = -1.1$  bin may be affected by a reddening of  $E(B - V) \sim 0.2$  (note that 6 more objects with  $\alpha_{\text{opt}} \sim 0$  and 4 with  $\alpha_{\text{opt}} < -1.25$  are not included in Table 4, due to our  $N \geq 10$  requirement, §3.7.2). The dust suppresses  $\nu L_{\nu}(\text{FUV})$  by a factor of two and five in the  $\alpha_{\text{opt}} = -0.7$  and  $\alpha_{\text{opt}} = -1.1$  bins, indicating significant reprocessing of  $L_{\text{bol}}$  by the dust.

The distribution of extinction values appears independent of the AGN luminosity for  $\log L_{\text{bol}} = 43 - 46$ , as the tail of the  $\nu L_{\nu}(\text{UV})/L_{\text{bH}\alpha}$  distribution remains largely unchanged with  $L_{\text{bH}\alpha}$  (fig. 15, left panel). However, for the quasars sample, Richards et al. (2003) find that 6% of the quasars appear to have  $E(B - V) \sim 0.13$ , while Hopkins et al. (2004) find that only 2% of the SDSS quasars have  $E(B - V) > 0.1$  and 1% have  $E(B - V) > 0.2$ , significantly lower fractions than found here (but note that some of it may be due to different procedures used to derive  $E(B - V)$ ). This suggests a decreasing reddening with increasing  $L_{\text{bol}}$ , which may be related to the observed drop in the mid-IR to  $L_{\text{bol}}$  with increasing  $L_{\text{bol}}$  in AGN (Maiolino et al. 2007; Treister et al. 2008) suggesting



**Figure 20.** The FUV, X-ray, and broad H $\alpha$  emission ratios as a function of  $m$  and  $l$  (left column), and as a function of  $L_{bH\alpha}$  and  $\Delta v$  (right column), at the  $L_{bH\alpha}$  luminosity range where the host contribution is negligible. All bins in each panel have the same number of objects (lower left corner). In search of the primary parameter which drives a given ratio, we select for the x-axis either  $m$  or  $l$  (left column), and  $L_{bH\alpha}$  or  $\Delta v$  (right column) to create the smallest spread for the different curves, where each curve includes T1 objects with a specific value of the other parameter. Error bars denote the error in the mean value in the bin. To reduce non-detection biases in the X-ray, in the two lower rows only  $F_{bH\alpha} > 10^{-13.5}$  erg s $^{-1}$  cm $^{-2}$  objects are used. (a,b) Note that the primary parameter which sets the mean  $vL_v(FUV)/L_{bH\alpha}$  is  $l$ . The ratio is independent of  $M_{BH}$ . Binning by  $L_{bH\alpha}$  and  $\Delta v$  produces a larger dispersion. (c,d) Here either  $m$  or  $L_{bH\alpha}$  are the dominant parameters. (e,f)  $\alpha_{FUV,X}$ , equivalent to the ratio of the upper two ratios. Here  $L_{bH\alpha}$  appears to be the dominant parameter.

a drop in the covering factor of dusty gas with increasing luminosity. However, the mid-IR emission may be associated with the ‘receding torus’ scenario (Lawrence & Elvis 1982) where the associated dust is X-ray absorbing and thus optically very thick. The dust associated with the obscuring torus will therefore not be the reddening dust. Also, as noted above, the T1 sample does not show a luminosity dependence for the reddened fraction of the objects.

An alternative interpretation of the lower fraction of reddened AGN in the Richard et al. and Hopkins et al. quasar samples, is a relation of reddening with  $L/L_{Edd}$ . As discussed above, quasars have a higher  $L/L_{Edd}$  by selection (Fig. 8) compared to emission line selected T1 sample, and higher  $L/L_{Edd}$  AGN tend to have lower reddening (Fig. 20).

What is the nature of the dust? Richards et al. (2003) and Hopkins et al. (2004) find that the SED of their red quasars is well reproduced, down to  $\sim 1500\text{\AA}$ , by extinction of SMC-like dust. A similar conclusion is reached here based on  $\alpha_{opt}$  (Fig. 17). The local UV slope argues against SMC-like dust, but given the large scatter

and the lack of spectroscopy in the UV, the discrepancy may not be significant.

Where is the dust located? The location of the reddening dust can be explored by comparing the near, mid and far-IR emission of reddened and unreddened objects, which can tell how far from the center the bulk of the reddening dust resides. The possible relation of the reddening with the host/AGN luminosity ratio, or (equivalently) with  $L/L_{Edd}$ , quantities which are not inclination dependent, suggests the reddening dust does not reside on the smallest possible scales.

### 4.3 The host galaxy

#### 4.3.1 $L_{host}$ vs. $L_{AGN}$

In §3.5.1, we use the  $z$  band host galaxy luminosity as a measure of the host stellar mass (within the SDSS aperture). We find that the correlation of the T1 host luminosity with the AGN luminosity ( $L_{bH\alpha}$ ) is induced by the correlation of both quantities with  $z$ .

Partial correlation analysis of the host luminosity versus  $L_{\text{bH}\alpha}$ , for a fixed  $z$ , yields no significant correlation. A similar result was obtained by Hao et al. (2005b) for emission line selected type 1 and type 2 AGN from the SDSS. It reflects, as noted by Hao et al., the large range of  $L/L_{\text{Edd}}$  in emission line selected AGN samples, in contrast with colour and quasi-stellar selected AGN samples, which by selection have a high  $L/L_{\text{Edd}}$  (e.g. §3.1.2), which induces a correlation between the AGN and the host luminosity.

#### 4.3.2 The host type

Kauffmann et al. (2003a, hereafter K03) compared the host properties of type 2 AGN to inactive galaxies in the SDSS survey, and found that type 2 AGN reside almost exclusively in massive galaxies. Here, we verify this effect, and find that at low  $z$  the mean  $z$  band host luminosity of type 1 and type 2 AGN, are a factor of three to five higher than in inactive galaxies. We further separated the inactive galaxies to NEG and SFG, and find that at the lowest  $z$ , the  $z$  band luminosity of T1 AGN hosts follow very closely the mean NEG luminosity. So, it appears that the preference of AGN for more massive hosts may just reflect the preference for NEG, and the fact that NEG tend to be more massive.

Further support for NEG as the host of the T1 AGN comes from a comparison of the host luminosity distributions. The NEG and the T1 hosts follow similar (fibre  $r$ -band) luminosity distribution. The relative normalization of T1 AGN is  $\sim 3\%$  of the NEG. This result is consistent with a scenario where  $\sim 3\%$  of NEG host a T1 AGN, irrespective of the NEG luminosity. The similar stellar mass distribution of NEG and of type 2 AGN is also noted in Salim et al. (2007, Fig. 17 there). In contrast, if T1 occur in SFG, then the fraction of SFG hosting a T1 AGN increases from  $\sim 10^{-3}$  to  $\sim 1$  over the observed range of luminosity (Fig. 9). Such a drastic change in the AGN fraction with the host luminosity appears implausible. Furthermore, it contradicts the fact that the T1 hosts are more massive than SFG (Fig. 12, see also Salim et al. 2007).

With increasing luminosity we find that the hosts of AGN become bluer, as was first noted for quasars (Hutchings 1987; Kirhakos et al. 1999; Jahnke et al. 2004). This was verified for type 2 AGN by K03 who found an increasing fraction of younger stars with increasing luminosity. A similar trend was noted by Vanden Berk et al. (2006) for type 1 SDSS AGN, although the non orthogonality of the host and AGN eigenvectors leads to a non-unique solution in the method used in that study (see §2.2). Here we find that the T1 and type 2 hosts are similar, and with increasing  $L_{\text{bH}\alpha}$ , the mean host  $u-z$  colour for both types gets bluer (Fig. 12). For  $L_{\text{bH}\alpha} < 10^{41}$  erg s $^{-1}$ , the T1 host colour is red, identical to that of NEG, and for  $L_{\text{bH}\alpha} > 10^{41}$  erg s $^{-1}$ , it becomes bluer, becoming intermediate between NEG and SFG at  $L_{\text{bH}\alpha} \gtrsim 3 \times 10^{42}$  erg s $^{-1}$ . This colour effect is also characterized by the tendency of AGN (depending on the sample selection criteria, e.g. Hickox et al. 2009) to reside in the ‘green valley’ in the colour-luminosity diagrams (e.g. Fig. 1 in Salim et al. 2007). The implied specific SFR ratios of  $10^{-2.9} < b_{300} < 10^{-2.25}$  (Table 3) are significantly lower than for ‘starburst’ galaxies, i.e. galaxies forming stars faster than their mean SFR over the Hubble time ( $\log b_{300} > -1.5$ ). The highest luminosity T1 AGN are intermediate between NEG and SFG, as found in type 2 AGN (Heckman et al. 2004; Wild et al. 2007). However, their concentration index and bulge/total light ratio (K03) remains similar to NEG, indicating early type hosts.

Why do low luminosity AGN avoid SFG? Is it due to suppression of the host SFR? This scenario is unlikely given the low AGN luminosity, which is well below the host luminosity, and thus

unlikely to affect the host significantly. In fact, as  $L_{\text{bol}}$  increases and the AGN dominates the host emission, the host SFR actually increases, the reverse of what is expected from the AGN SFR suppression scenario. Furthermore, the suppression scenario is inconsistent with the various AGN host galaxy parameters (stellar mass, concentration index, bulge/total light ratio) which indicate AGN occur in bulge dominated galaxies (see K03), in contrast with the average galaxy parameters of SFG which characterize disc dominated galaxies (Kauffmann et al. 2003b). A similar preference for earlier type hosts was also found by Ho et al. (1997b) for the lowest luminosity AGN ( $L_{\text{bH}\alpha} \sim 10^{39}$  erg s $^{-1}$ ), for luminous quasars (e.g. McLeod & Rieke 1995), and for X-ray selected AGN (e.g. Schade et al. 2000). The early type host galaxy parameters also argue against a delay scenario, where the host SF episode shuts off by the time the infalling gas reaches the AGN. Massive stars will indeed die off within  $\sim 100$  Myr, but the structure of the galaxy will take longer to evolve. Since low luminosity AGN and SFG appear to have different host mass distribution properties, they are not likely to be different phases in the same host population.

An alternative explanation is that AGN avoid SFG because SFG tend to have lower mass bulges. If SFG at low  $z$  typically have  $M_{\text{BH}} \ll 10^6 M_{\odot}$ , yet rather luminous discs, the AGN may be optically swamped by the host, even if it shines close to  $L_{\text{Edd}}$ . Another option is obscuration by dust, as high SFR is generally associated with excess dust extinction. For example, Simões Lopes et al. (2007) find excess dust even in early type galaxies with AGN activity. To test whether AGN avoid SFG because of their high SFR and the associated dust, or because of their low bulge mass, one needs to compare the fraction of SFG and the fraction of NEG galaxies which host an AGN, at the same bulge mass, based on X-ray and mid-IR surveys which are less affected by dust extinction. If the fraction is similar, then the SF activity has no effect on the AGN activity.

One should note that the B04 definition of an inactive galaxy as SFG or NEG is not physically motivated, but is just based on the ability to detect and measure a set of emission lines. It will clearly be useful to compare the T1 host properties with inactive galaxies separated based on a physical property, such as the absolute or specific SFR.

#### 4.3.3 The SFR values

The  $u$  band luminosity is mostly set by the SFR in the recent  $\sim 100$  Myr. At  $L_{\text{bol}} \sim 5 \times 10^{44}$  erg s $^{-1}$  the T1 hosts  $u$  band luminosity is intermediate between SFG and NEG (Fig. 12), which suggests an intermediate SFR. A similar intermediate SFR was found by Salim et al. (2007, Fig. 3 there) for the SDSS type 2 AGN, based on the UV luminosity from GALEX (see also Heckman et al. 2004). The value of the SFR can be estimated from the  $u$  band luminosity (Moustakas et al. 2006, eq. 11). The mean host  $u$  band luminosity of the T1 AGN then implies a SFR which increases with the AGN  $L_{\text{bol}}$ . The SFR increases from  $0.11 M_{\odot} \text{ yr}^{-1}$  at  $\log L_{\text{bol}} = 42.2$  to  $3.5 M_{\odot} \text{ yr}^{-1}$  at  $\log L_{\text{bol}} = 44.6$ , with a mean SFR of  $\sim 1.1 M_{\odot} \text{ yr}^{-1}$  for the most abundant  $\log L_{\text{bol}}$  of 43.7 (derived from  $L_{\text{bH}\alpha} = 41.6$ , the most abundant T1 AGN, Fig. 5). The rise in the host SFR with  $L_{\text{bol}}$  is qualitatively consistent with earlier results for type 2 AGN (K03, Silverman et al. 2009, Netzer 2009). The values derived above are comparable to the SFR vs.  $L_{\text{bol}}$  values presented by Trakhtenbrot & Netzer (2010, Fig. 1 there) for SDSS type 2 AGN, where the SFR is derived from the  $D_{4000}$  (B04), and  $L_{\text{bol}}$  is derived from the [O III] and [O I] luminosities. Thus, we conclude that inferring  $L_{\text{bol}}$  from either the NLR ([O III]) or the BLR (H $\alpha$ ),

and inferring the SFR from either the  $u$  band luminosity or from  $D_{4000}$ , leads to similar  $L_{\text{bol}}$  versus SFR relations for type 2 and type 1 AGN.

An estimate of the SFR in type 1 AGN was made by Kim et al. (2006), based on the [O II] luminosity, assuming 1/3 comes from star formation, and 2/3 is powered by the AGN. They used the Kewley et al. (2004) [O II] calibration of the SFR and derived a SFR for the most abundant  $L_{[\text{O II}]} = 5 \times 10^{40} \text{ erg s}^{-1}$  of  $0.5 \text{ M}_{\odot} \text{ yr}^{-1}$ , for an SDSS based sample of  $z < 0.3$  type 1 AGN, which should overlap well the T1 sample. Kim et al. note the [O II] emission may all be powered by the AGN, so the actual SFR could be even smaller. Note that the [O II] based SFR values for the SDSS type 2 AGN, presented by Silverman et al. (2009, Fig. 10 there), is typically in the range of  $0.1 - 1 \text{ M}_{\odot} \text{ yr}^{-1}$ , which is significantly below the UV and the H $\alpha$  based estimate for type 2 SDSS AGN, of  $1 - 10 \text{ M}_{\odot} \text{ yr}^{-1}$  (Salim et al. 2007, Fig. 3 there). So, apparently there is a discrepancy between the  $u/UV$  and the [O II] derived SFR in type 2 AGN (see discussion in Ho 2005), which may also be present in the T1 sample, if [O II] is mostly powered by the AGN, and the implied SFR should be lower. Since [O II] measures the most recent SFR compared to the UV/ $u$  continuum, the discrepancy may imply a post starburst.

## 5 CONCLUSION

We present and analyze a new sample (T1) of 3579 broad H $\alpha$  selected AGN from the SDSS DR7, with  $\log L_{\text{bH}\alpha} = 40 - 44$ , which spans  $m = 6 - 9$  and  $l = -3 - 0$ . We add UV (GALEX), IR (2MASS), and X-ray (ROSAT) luminosities to form the mean SED. The main results are:

- (i) The H $\alpha$  FWHM velocity distribution  $dN/d\log \Delta v$  is independent of luminosity and falls exponentially with  $\Delta v$ . The origin of this distribution remains to be understood.
- (ii) The observed mean 9000Å–1500Å SED, as a function of  $L_{\text{bH}\alpha}$ , is consistent with a sum of the mean SED of luminous quasars, which scales linearly with  $L_{\text{bH}\alpha}$ , and a host galaxy contribution.
- (iii) The host galaxy  $r$ -band luminosity function of T1 objects with an extended morphology follows the NEG luminosity function, with a relative normalization of  $\sim 3\%$ , suggesting that the host of broad line AGN are NEG, and the AGN probability of occurrence is independent of the host mass.
- (iv) The mean  $z$  band luminosity and the  $u - z$  band colour of the lowest luminosity T1 host is identical to NEG. The host colour becomes progressively bluer with increasing luminosity. The implied mean SFR versus  $L_{\text{bol}}$  is similar to that found in type 2 AGN.
- (v) The dispersion in the optical-UV SED in luminous AGN ( $\log L_{\text{bH}\alpha} \geq 43$ ), is consistent with reddening. This indicates the intrinsic SED of AGN is blue, with a small dispersion, as predicted from thermal thin accretion disc models.
- (vi) Reddening by dust along the line of sight to T1 AGN is common (40% with  $E(B - V) \gtrsim 0.1$ ).
- (vii) The  $L_{\text{bH}\alpha}$  versus  $L_X$  correlation provides a useful probe for unobscured narrow line AGN. It can be used to test if the absence of a broad H $\alpha$ , in X-ray detected AGN, is significant.
- (viii) The primary parameter which drives  $\alpha_{\text{FUV},X}$  is the luminosity, rather than  $M_{\text{BH}}$  or  $L/L_{\text{Edd}}$ .
- (ix) The primary parameter which drives  $\log vL_{\nu}(\text{FUV})/\log L_{\text{bH}\alpha}$  is  $L/L_{\text{Edd}}$ , which may indicate that lower  $L/L_{\text{Edd}}$  AGN are more likely to be reddened.

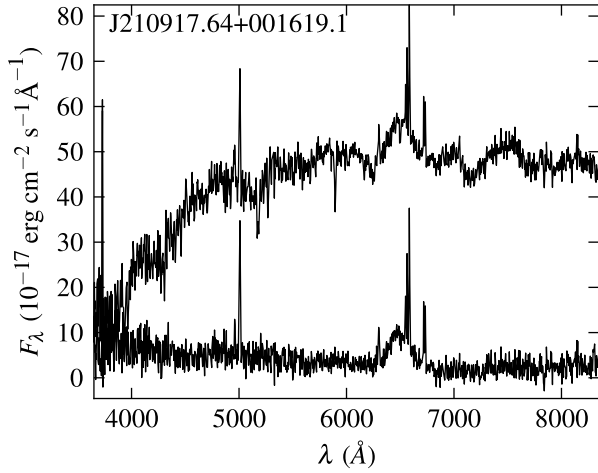
This publication makes use of data products from the SDSS project, funded by the Alfred P. Sloan Foundation, data from GALEX supported by NASA, data from the Two Micron All Sky Survey, funded by the NASA and the NSF, and from the ROSAT Data Archive of the Max-Planck-Institut für extraterrestrische Physik (MPE) at Garching, Germany. A.L. acknowledges support by the Israel Science Foundation grant 407/08.

## REFERENCES

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2004, AJ, 128, 502
- Abazajian, K. N., et al. 2009, ApJS, 182, 543
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
- Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54
- Baskin, A., & Laor, A. 2005, MNRAS, 356, 1029
- Bauer, F. E., Alexander, D. M., Brandt, W. N., Schneider, D. P., Treister, E., Hornschemeier, A. E., & Garmire, G. P. 2004, AJ, 128, 2048
- Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
- Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. 2009, ApJ, 697, 160
- Bica, E., & Alloin, D. 1986, A&A, 162, 21
- Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734
- Bonning, E. W., Cheng, L., Shields, G. A., Salvander, S., & Gebhardt, K. 2007, ApJ, 659, 211
- Boroson, T. A., & Green, R. F. 1992, ApJS, 80, 109 (BG92)
- Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, 351, 1151 (B04)
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Caccianiga, A., & Severgnini, P. 2011, MNRAS, 415, 1928
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
- Cutri, R. M., et al. 2003, Explanatory Supplement to the 2MASS All Sky Data Release (Washington: NASA), <http://www.ipac.caltech.edu/2mass/releases/all-sky/doc/explsup.html>
- Croom, S. M., et al. 2002, MNRAS, 337, 275
- Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, MNRAS, 349, 1397
- Croom, S. M. 2011, ApJ, 736, 161
- Davis, S. W., Woo, J.-H., & Blaes, O. M. 2007, ApJ, 668, 682
- Davis, S. W., & Laor, A. 2011, ApJ, 728, 98
- Dietrich, M., Hamann, F., Shields, J. C., Constantin, A., Vestergaard, M., Chaffee, F., Foltz, C. B., & Junkkarinen, V. T. 2002, ApJ, 581, 912
- Elitzur, M., & Shlosman, I. 2006, ApJ, 648, L101
- Filippenko, A. V. 1997, ARA&A, 35, 309
- Fine, S., Croom, S. M., Hopkins, P. F., et al. 2008, MNRAS, 390, 1413
- Fine, S., Croom, S. M., Bland-Hawthorn, J., et al. 2010, MNRAS, 409, 591
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
- Gaskell, C. M., Goosmann, R. W., Antonucci, R. R. J., & Whysong, D. H. 2004, ApJ, 616, 147



- Gaskell, C. M., & Benker, A. J. 2007, arXiv:0711.1013
- Green, R. F., Schmidt, M., & Liebert, J. 1986, *ApJS*, 61, 305
- Greene, J. E., & Ho, L. C. 2004, *ApJ*, 610, 722
- Greene, J. E., & Ho, L. C. 2005, *ApJ*, 630, 122
- Greene, J. E., & Ho, L. C. 2007, *ApJ*, 667, 131 (GH07)
- Grupe, D., Komossa, S., Leighly, K. M., & Page, K. L. 2010, *ApJS*, 187, 64
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- Hao, L., et al. 2005a, *AJ*, 129, 1783
- Hao, L., Strauss, M. A., Fan, X., et al. 2005b, *AJ*, 129, 1795
- Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, *ApJ*, 613, 109
- Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1995, *AJ*, 109, 1498
- Hickox, R. C., et al. 2009, *ApJ*, 696, 891
- Ho, L. C., Filippenko, A. V., Sargent, W. L. W., & Peng, C. Y. 1997a, *ApJS*, 112, 391
- Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997b, *ApJ*, 487, 568
- Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 2003, *ApJ*, 583, 159
- Ho, L. C. 2005, *ApJ*, 629, 680
- Hopkins, P. F., et al. 2004, *AJ*, 128, 1112
- Hutchings, J. B. 1987, *ApJ*, 320, 122
- Jahnke, K., Kuhlbrodt, B., & Wisotzki, L. 2004, *MNRAS*, 352, 399
- Just, D. W., Brandt, W. N., Shemmer, O., et al. 2007, *ApJ*, 665, 1004
- Kaspi, S., Maoz, D., Netzer, H., Peterson, B. M., Vestergaard, M., & Jannuzi, B. T. 2005, *ApJ*, 629, 61
- Kauffmann, G., et al. 2003a, *MNRAS*, 346, 1055 (K03)
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, *MNRAS*, 341, 54
- Kelly, B. C., Bechtold, J., Trump, J. R., Vestergaard, M., & Siemiginowska, A. 2008, *ApJs*, 176, 355
- Kewley, L. J., Geller, M. J., & Jansen, R. A. 2004, *AJ*, 127, 2002
- Kim, M., Ho, L. C., & Im, M. 2006, *ApJ*, 642, 702
- Kirhakos, S., Bahcall, J. N., Schneider, D. P., & Kristian, J. 1999, *ApJ*, 520, 67
- Kollmeier, J. A., et al. 2006, *ApJ*, 648, 128 (K06)
- Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1994, *ApJ*, 435, 611
- Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1997, *ApJ*, 477, 93
- Laor, A. 2003, *ApJ*, 590, 86
- Laor, A., & Davis, S. W. 2011, *MNRAS*, 417, 681
- Magorrian, J., et al. 1998, *AJ*, 115, 2285
- Lawrence, A., & Elvis, M. 1982, *ApJ*, 256, 410
- Maiolino, R., Shemmer, O., Imanishi, M., Netzer, H., Oliva, E., Lutz, D., & Sturm, E. 2007, *A&A*, 468, 979
- Maoz, D., Nagar, N. M., Falcke, H., & Wilson, A. S. 2005, *ApJ*, 625, 699
- Martin, D. C., et al. 2005, *ApJ*, 619, L1
- McLeod, K. K., & Rieke, G. H. 1995, *ApJ*, 454, L77
- Morrissey, P., et al. 2007, *ApJS*, 173, 682
- Moustakas, J., Kennicutt, R. C., Jr., & Tremonti, C. A. 2006, *ApJ*, 642, 775
- Netzer, H., Shemmer, O., Maiolino, R., et al. 2004, *ApJ*, 614, 558
- Netzer, H. 2009, *MNRAS*, 399, 1907
- Neugebauer, G., Green, R. F., Matthews, K., Schmidt, M., Soifer, B. T., & Bennett, J. 1987, *ApJS*, 63, 615
- Nicastro, F. 2000, *ApJ*, 530, L65
- Osterbrock, D. E. 1981, *ApJ*, 249, 462
- Pei, Y. C. 1992, *ApJ*, 395, 130
- Pfeffermann, E., et al. 1987, *Proc. SPIE*, 733, 519
- Phillips, M. M. 1978, *ApJS*, 38, 187
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press, —c1992, 2nd ed.,
- Richards, G. T., et al. 2003, *AJ*, 126, 1131
- Richards, G. T., et al. 2006, *ApJS*, 166, 470 (R06)
- Rigby, J. R., Rieke, G. H., Donley, J. L., Alonso-Herrero, A., & Pérez-González, P. G. 2006, *ApJ*, 645, 115
- Salim, S., Rich, R. M., Charlot, S., et al. 2007, *ApJS*, 173, 267
- Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T., & Matthews, K. 1989, *ApJ*, 347, 29
- Schade, D. J., Boyle, B. J., & Letawsky, M. 2000, *MNRAS*, 315, 498
- Schartel, N., et al. 1996, *MNRAS*, 283, 1015
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schneider, D. P., Richards, G. T., Fan, X., et al. 2002, *AJ*, 123, 567
- Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2005, *AJ*, 130, 367
- Schneider, D. P., et al. 2010, *AJ*, 139, 2360
- Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, *ApJ*, 680, 169
- Shen, Y., et al. 2011, *ApJS*, 194, 45
- Shi, Y., Rieke, G. H., Smith, P., Rigby, J., Hines, D., Donley, J., Schmidt, G., & Diamond-Stanic, A. M. 2010, *ApJ*, 714, 115
- Silverman, J. D., Lamareille, F., Maier, C., et al. 2009, *ApJ*, 696, 396
- Simões Lopes, R. D., Storchi-Bergmann, T., de Fátima Saraiva, M., & Martini, P. 2007, *ApJ*, 655, 718
- Skrutskie, M. F., et al. 2006, *AJ*, 131, 1163
- Soltan, A. 1982, *MNRAS*, 200, 115
- Stark, A. A., Gammie, C. F., Wilson, R. W., Bally, J., Linke, R. A., Heiles, C., & Hurwitz, M. 1992, *ApJS*, 79, 77
- Steffen, A. T., Strateva, I., Brandt, W. N., Alexander, D. M., Koekemoer, A. M., Lehmer, B. D., Schneider, D. P., & Vignali, C. 2006, *AJ*, 131, 2826
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Strauss, M. A., et al. 2002, *AJ*, 124, 1810
- Trammell, G. B., Vanden Berk, D. E., Schneider, D. P., Richards, G. T., Hall, P. B., Anderson, S. F., & Brinkmann, J. 2007, *AJ*, 133, 1780
- Tran, H. D., Lyke, J. E., & Mader, J. A. 2011, *ApJ*, 726, L21
- Trakhtenbrot, B., & Netzer, H. 2010, *MNRAS*, 406, L35
- Treister, E., Krolik, J. H., & Dullemond, C. 2008, *ApJ*, 679, 140
- Trump, J. R., et al. 2009, *ApJ*, 706, 797
- Vanden Berk, D. E., et al. 2001, *AJ*, 122, 549
- Vanden Berk, D. E., et al. 2006, *AJ*, 131, 84
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525
- Vasudevan, R. V., & Fabian, A. C. 2009, *MNRAS*, 392, 1124
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, 63, 295
- Voges, W., et al. 1999, *A&A*, 349, 389
- Voges, W., et al. 2000, *IAU Circ.* 7432, 3
- Wang, T.-G., Dong, X.-B., Zhang, X.-G., Zhou, H.-Y., Wang, J.-X., & Lu, Y.-J. 2005, *ApJ*, 625, L35
- Ward, M., Elvis, M., Fabbiano, G., et al. 1987, *ApJ*, 315, 74
- Wild, V., Kauffmann, G., Heckman, T., et al. 2007, *MNRAS*, 381, 543
- Wyder, T. K., et al. 2007, *ApJS*, 173, 293
- Yip, C. W., et al. 2004, *AJ*, 128, 585 (Yip04)



**Figure A1.** An example of host subtraction using the Yip04 eigenspectra (§2.2). The top line is the SDSS spectrum, after correcting for Galactic extinction. The bottom line is the eigenspectra-subtracted spectrum. Note the high contrast of the broad H $\alpha$  in the residual.

Zheng, W., Kriss, G. A., Telfer, R. C., Grimes, J. P., & Davidsen, A. F. 1997, *ApJ*, 475, 469

## Appendices

### APPENDIX A: SAMPLE CREATION DETAILS

#### A1 Bad pixels

During the initial filtering of the SDSS database (§2.1), we require each spectrum to comply with the following criteria, concerning ‘bad’ pixels. A pixel is considered as bad if it has any of the flags listed in table 11 of Stoughton et al. (2002), except ‘MANYBAD-COLUMNS’ and ‘NEARBADPIXEL’. Also, due to SDSS over-sampling, each pixel within two pixels of a flagged pixel is considered bad.

- < 20% bad pixels near the broad H $\alpha$  (6250Å–6880Å)
- 0 bad pixels near the narrow H $\alpha$  (6530Å–6600Å)
- < 50% bad pixels in the continuum fitting ranges (6125Å–6250Å & 6880Å–7000Å)

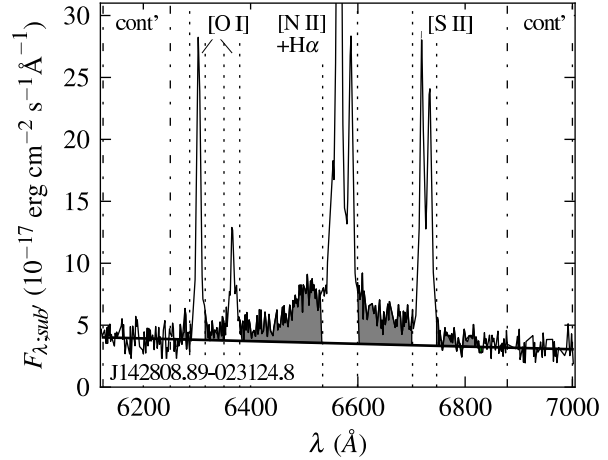
#### A2 Host subtraction

In §2.2, each spectrum is fit with three galaxy eigenspectra (Yip et al. 2004) and a power-law component. The fit is performed on the following wavelength ranges, chosen since they are devoid of the strong AGN and stellar emission lines: 3455Å–3490Å, 3500Å–3550Å, 3600Å–3700Å, 4200Å–4250Å, 4400Å–4460Å, 4490Å–4600Å, 5130Å–5680Å, 5750Å–5830Å, 5920Å–6050Å, and 6130Å–6200Å.

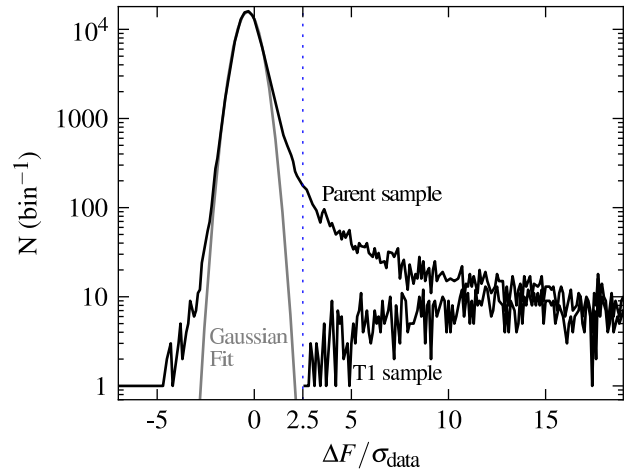
An example of a spectrum before and after subtracting the fit of the three eigenspectra is shown in Figure A1.

#### A3 Broad line flux

A depiction of the excess flux identification process presented in §2.3 is shown in Figure A2.



**Figure A2.** Identifying broad H $\alpha$  emission (§2.3), shown on an example host-subtracted spectrum. The continuum (solid straight line) is interpolated from the mean of the regions bordered by dash-dot lines. The potential BLR flux ( $\Delta F$ ) is filled with gray. Flux near the strong narrow lines (delimited by dotted lines) does not enter the calculation of  $\Delta F$ .



**Figure A3.** The distribution of the normalized excess flux (§2.3) in the parent sample (232 837 objects). Normalization is by the  $F_\lambda$  dispersion ( $\sigma_{\text{data}}$ ) in the continuum windows (Fig. A2). Bin size is 0.1. To avoid objects with excess flux due to the normal dispersion (see Gaussian fit to the distribution), only the 6 986 objects with  $\Delta F/\sigma_{\text{data}} > 2.5$  passed to the next stage of the algorithm. Also shown is the final T1 sample, the 3 579 objects that pass the final criteria in §2.5.

Figure A3 shows the  $\Delta F/\sigma_{\text{data}}$  histogram of the parent sample (§2.1). Near  $\Delta F/\sigma_{\text{data}} \approx 0$ , the distribution fits a Gaussian, as expected from the central limit theorem. The objects in the positive wing are potential T1 objects. We choose a criterion of  $\Delta F/\sigma_{\text{data}} > 2.5$ , since objects with lower  $\Delta F/\sigma_{\text{data}}$  have a < 1% chance to pass the additional criteria described in §2.5. The  $\Delta F/\sigma_{\text{data}}$  distribution of the T1 sample, after the criteria in §2.5, is also shown. It can be seen that a criterion of  $\Delta F/\sigma_{\text{data}} \gtrsim 12$  would have been roughly sufficient to avoid false identifications.

#### A4 Broad H $\alpha$ and narrow lines fit

The emission lines are fit with Gauss-Hermite functions (GHs) in three stages (§2.4). The details of this process are described below. We use the conventions and notations of van der Marel & Franx (1993):  $\gamma, V, \sigma$  and  $h^i$  are the flux, mean velocity, dispersion, and  $i^{\text{th}}$  Hermite coefficient, respectively. Fourth-order GHs have finite  $h^3$  and  $h^4$ , while the other  $h^i$  are zero. A line name in subscript denotes a property of a specific line. The SDSS flux density error at  $\lambda$  is denoted by  $\varepsilon_\lambda$ .

##### A4.1 [O III]

If there are  $< 20\%$  bad pixels at  $4967\text{\AA}$ – $5250\text{\AA}$ , the range is fit with three components:

- An iron template derived from observations of I Zw 1 (kindly provided by T. Boroson). We fit its flux, velocity, and width (by convolving with a Gaussian), denoted  $\gamma_{\text{Fe II}}, V_{\text{Fe II}}$  and  $\sigma_{\text{Fe II}}$ , respectively. The value of  $\gamma_{\text{Fe II}}$  is initialized from the power-law coefficient found in §2.2, assuming a Fe II equivalent width of  $7\text{\AA}$  in the fit range. During the fit we keep  $|V_{\text{Fe II}}| < 550 \text{ km s}^{-1}$  and  $\sigma_{\text{Fe II}} < 20000 \text{ km s}^{-1}$ .
- A broken power law for the underlying continuum and the red wing of the broad H $\beta$ . Initialization is as follows: the wavelength and flux at the break –  $5082\text{\AA}$ ,  $F_\lambda(5082\text{\AA})$ ; shorter  $\lambda$  slope – interpolation between  $F_\lambda(4967\text{\AA})$  and  $F_\lambda(5082\text{\AA})$ ; longer  $\lambda$  slope – 0.
- A 4<sup>th</sup>-order GH for [O III]. The  $\gamma_{[\text{O III}]}, V_{[\text{O III}]}$  and  $\sigma_{[\text{O III}]}$  are initialized from the SDSS fit. The values of  $h_{[\text{O III}]}^3$  and  $h_{[\text{O III}]}^4$  are initialized to 0. During the fit we keep  $|V_{[\text{O III}]}| < 550 \text{ km s}^{-1}$  and  $\text{FWHM}_{[\text{O III}]} < 1400 \text{ km s}^{-1}$ . The fit is run once with  $h_{[\text{O III}]}^3$  and  $h_{[\text{O III}]}^4$  fixed to 0 (a Gaussian), and once when they are allowed to vary. The final fit is chosen according to an F-test ( $p < 0.05$ ).

The fit is disregarded if there are bad pixels  $< 6\text{\AA}$  from the fit [O III] peak. The [O III] is considered ‘detected’ if it has  $F_\lambda > 3.5 \times \varepsilon_{[\text{O III}]}$ , and  $V_{[\text{O III}]}$  has not reached the limit mentioned above.

##### A4.2 H $\alpha$ region

We use nine components to fit the  $6250\text{\AA}$ – $6880\text{\AA}$  region:

- A continuum.
- An up to 10<sup>th</sup>-order GH for the broad H $\alpha$  (bH $\alpha$ ).
- Seven up to 4<sup>th</sup>-order GHs for the narrow lines (NLs). If no good pixel is found within  $5\text{\AA}$  of the rest wavelength of a NL, it is not fit. Below, a subscript ‘NL’ refers to all seven NLs simultaneously.

Initially, the broad H $\alpha$  is fit by a 4<sup>th</sup>-order GH, in order to avoid fitting the narrow lines by high-order Hermite polynomials. The mean velocities of NLs in a doublet are always kept equal, and the flux ratios in the [N II] and [O I] doublets are fixed to 3. We limit  $-550 < V_{\text{NL}} < 550 \text{ km s}^{-1}$ . The Levenberg-Marquardt best-fitting algorithm (Press et al. 1992) is used.

Several fit attempts are made, differing by the initialization and attainable values of the parameters:

- A.** Initialization of the  $\gamma_{\text{bH}\alpha}$ ,  $\sigma_{\text{bH}\alpha}$  and continuum is based on the continuum and excess flux found in §2.3. The  $V_{\text{bH}\alpha}$  and all  $h_{\text{bH}\alpha}^i$  are initialized to 0. The  $V_{\text{NL}}, \sigma_{\text{NL}}, h_{\text{NL}}^3$  and  $h_{\text{NL}}^4$  are initialized by the fit properties of [O III]. The  $\gamma_{\text{NL}}$  are initialized

using NL flux ratios of typical Seyferts<sup>10, 11</sup>. During the fit,  $\sigma_{\text{NL}}, h_{\text{NL}}^3$  and  $h_{\text{NL}}^4$  are kept fixed to the [O III] values.

- B.** Similar to **A**, but starting from the result of **A** and allowing all  $\sigma_{\text{NL}}$  to vary together (one additional degree of freedom). We limit  $\sigma_{\text{NL}} < 1.5 \times \sigma_{[\text{O III}]}$ <sup>12</sup>.
- C.** In this attempt we avoid implausibly large troughs in the broad H $\alpha$  profile, which can be created by the [N II] and nH $\alpha$  components. The maximum  $F_\lambda$  of these 3 NLs are limited such that the broad profile will not have a trough deeper than 20% of its local  $F_\lambda$ . Also, the error entering the fit near these NLs ( $\pm 10\text{\AA}$ ) is reduced to a 1/3 of its initial value, effectively improving the fit of this region on the expense of regions farther from H $\alpha$ . The bH $\alpha$  and continuum are initialized as in **A**. The  $\gamma_{\text{NL}}, V_{\text{NL}}$  and  $\sigma_{\text{NL}}$  are initialized from the result of **B**, keeping  $\sigma_{\text{NL}} < 1.5 \times \sigma_{[\text{O III}]}$  during the fit. The  $h_{\text{NL}}^3, h_{\text{NL}}^4$  are initialized to 0, and their absolute value is kept below  $h_{[\text{O III}]}^3$  and  $h_{[\text{O III}]}^4$ , respectively. During the fit,  $\sigma_{\text{NL}}, h_{\text{NL}}^3$  and  $h_{\text{NL}}^4$  of different NLs are kept equal.
- D.** The bH $\alpha$  and continuum are initialized as in **A**. The  $\gamma_{\text{NL}}$  and  $V_{\text{NL}}$  are initialized from the highest- $F_\lambda$  pixel within  $5\text{\AA}$  of the rest wavelength of the line. The  $\sigma_{\text{NL}}$  is initialized so that  $\text{FWHM}_{\text{NL}} = 300 \text{ km s}^{-1}$  and limited to  $< 1.5 \times \sigma_{[\text{O III}]}$ . As before, all  $\sigma_{\text{NL}}$  vary together. The  $h_{\text{NL}}^3$  and  $h_{\text{NL}}^4$  are fixed to 0.

The  $\chi^2$  of the **B**, **C** and **D** fits are compared, using the sum of the SDSS flux density error and 10% of the fit flux density as the error, and the best result is chosen. Objects in which  $\sigma_{\text{nH}\alpha}$  reached the limit ( $1.5 \times \sigma_{[\text{O III}]}$ ) are found to have non- or barely-detectable NLs near H $\alpha$ . When the narrow H $\alpha$  component is not restricted by other NLs, it erroneously fits the top of the broad H $\alpha$ . In such cases, we use result **A**, in which  $\sigma_{\text{NL}} = \sigma_{[\text{O III}]}$ . In the 45 of these objects that were fit with  $F_{\text{nH}\alpha}/F_{[\text{O III}]} > 2$ , the [O III] is weak as well, and the problem remains. Therefore, in these 45 objects we assume  $F_{\text{nH}\alpha}, F_{[\text{N II}]} = 0$ .

Finally, in order to fully account for the diversity of the broad H $\alpha$  profile, we run fits with a high-order GH for the broad H $\alpha$ . The initial values of the parameters are taken from the chosen fit above, while the new  $h^i$  are initialized to 0. Only the parameters of the broad H $\alpha$  and the  $\gamma_{\text{NL}}$ ’s vary during the fit. First, a fit is attempted with a 6<sup>th</sup>-order GH, using  $h_1$  &  $h_2$ <sup>13</sup>. Then we add  $h_5$  &  $h_6$ , and lastly  $h_7$  &  $h_8$ . Additional Hermite coefficients are retained only if they pass an F-test ( $p < 0.05$ ). To avoid having negative or distinct features in the broad H $\alpha$  profile, the broad H $\alpha$  flux density is set to zero at velocities beyond a velocity in which it is  $< 10\%$  of the flux density error.

After this process, 209 objects (3%) still have a reduced  $\chi^2 > 2$ . By eye-inspection, 80 of them have a broad H $\alpha$  and a reasonable fit. They were added to the T1 sample. Also, 35 show a BLR which was improperly fit. Since their number is relatively small, we do not attempt any further improvement of their fit, and they do not

<sup>10</sup>  $\gamma_{[\text{N II}] 6583\text{\AA}}, \gamma_{\text{H}\alpha}, \gamma_{[\text{O I}] 6300\text{\AA}}, \gamma_{[\text{S II}] 6731\text{\AA}}, \gamma_{[\text{S II}] 6716\text{\AA}}$  are initialized to 1/3, 1/3, 1/20, 1/9, 1/9 of  $\gamma_{[\text{O III}]}$ , respectively.

<sup>11</sup> If [O III] is not detected (0.6% of T1 sample),  $\gamma_{\text{NL}}$  initialization is derived from the upper limit on  $\gamma_{[\text{O III}]}$ . If bad pixels prevented fitting [O III] (4% of T1 objects),  $\gamma_{\text{NL}}$  and  $V_{\text{NL}}$  are initialized from the SDSS fits. In both cases  $\sigma_{\text{NL}}, h_{\text{NL}}^3$  and  $h_{\text{NL}}^4$  are initialized to a net Gaussian with  $\text{FWHM} = 300 \text{ km s}^{-1}$ .

<sup>12</sup> If [O III] is not detected, **B** is not run. Instead, **A** is run with  $\sigma_{\text{NL}}$  allowed to vary, keeping  $\text{FWHM}_{\text{NL}} < 650 \text{ km s}^{-1}$ .

<sup>13</sup> When  $V$  and  $\gamma$  of a GH are allowed to vary,  $h_1$  &  $h_2$  hold higher-order information than  $h^3$  &  $h^4$  (van der Marel & Franx 1993).

Object Name	$\Delta F / \sigma_{\text{data}}$	Rejection reason
J000202.71-010508.8	2.5	$\Delta v_{1000} > 25$
J000500.03+002055.1	2.5	$\Delta v_{1000} > 25$
J000605.59-092007.0	2.9	$\Delta F' / \sigma_{\text{data}} < 4$
J000834.72+003156.1	11.3	$\Delta F' / \sigma_{\text{data}} < 4$
J000911.58-003654.7	14.2	$F_{\lambda, \text{bH}\alpha} / \epsilon_{\lambda} < 2$

**Table B1.** Objects with excess flux near  $\text{H}\alpha$  that were rejected from the T1 sample. Col. 2 lists the normalized excess flux found in §2.3. The criterion which the object failed (§2.5) is listed in Col. 3. The electronic version of the paper includes all 3 407 objects.

enter the T1 sample. The rest of the high  $\chi^2$  objects do not show a clear broad  $\text{H}\alpha$ .

#### A4.3 $\text{H}\beta$

Finally, we fit the region  $\pm 6 \sigma_{\text{nH}\alpha}$  from the  $\text{H}\beta$  peak, using:

- A 4<sup>th</sup>-order GH for the narrow  $\text{H}\beta$ . Only  $\gamma_{\text{nH}\beta}$  is fit, initialized to  $\gamma_{\text{nH}\alpha}/3$ . Other GH parameters equal those fit to  $\text{H}\alpha$ .
- A parabola for the continuum + top of the broad  $\text{H}\beta$ . Initialization is by the edges and middle of the fit range. The non-linear coefficient of the parabola is non-zero only if necessary (F-test,  $p < 0.05$ ).

The fit is not performed if there are bad pixels less than  $4 \times \sigma_{\text{nH}\alpha}$  from the peak.

## APPENDIX B: REJECTED OBJECTS

Table B1 lists objects which passed the selection criteria described in §2.1 and §2.3, but failed one of the selection criteria described in §2.5.